```
OM nucleic - nucleic search, using sw model

Run on:         October  3, 2003, 11:40:01 ; Search time 8358.86 Seconds
                                             (without alignments)
                                             12307.996 Million cell updates/sec

Title:          US-10-014-717-1_COPY_11872_16104
Perfect score:  4233
Sequence:       1 atgacgatcaatcagcttct..........agggcaggagacgtagctaa 4233

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters:    45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      EST:*
                1:  em_estba:*
                2:  em_esthum:*
                3:  em_estin:*
                4:  em_estmu:*
                5:  em_estov:*
                6:  em_estpl:*
                7:  em_estro:*
                8:  em_htc:*
                9:  gb_est1:*
                10:  gb_est2:*
                11:  gb_htc:*
                12:  gb_est3:*
                13:  gb_est4:*
                14:  gb_est5:*
                15:  em_estfun:*
                16:  em_estom:*
                17:  em_gss_hum:*
                18:  em_gss_inv:*
                19:  em_gss_pln:*
                20:  em_gss_vrt:*
                21:  em_gss_fun:*
                22:  em_gss_mam:*
                23:  em_gss_mus:*
                24:  em_gss_pro:*
                25:  em_gss_rod:*
                26:  em_gss_phg:*
                27:  em_gss_vrl:*
                28:  gb_gss1:*
                29:  gb_gss2:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	175.4	4.1	812	29	BZ671349	BZ671349 PUBKS59TD
	2	174.4	4.1	627	29	BZ563956	BZ563956 pacs2-164
c	3	157.6	3.7	1552	29	BZ572480	BZ572480 msh2_2654
c	4	133	3.1	765	29	BZ562842	BZ562842 pacs2-164
c	5	128.8	3.0	646	28	AQ989512	AQ989512 Rfc00062
	6	125	3.0	1117	29	BZ548527	BZ548527 pacs1-60_
	7	121.6	2.9	921	29	BZ570019	BZ570019 msh2_1128
	8	120.8	2.9	1038	29	BZ567332	BZ567332 pacs2-164
	9	119	2.8	667	29	BZ548604	BZ548604 pacs1-60_
c	10	117	2.8	753	29	BZ561650	BZ561650 pacs2-164
	11	114.4	2.7	840	29	BZ558891	BZ558891 pacs2-164
	12	114	2.7	827	29	BZ680689	BZ680689 PUBEL33TD
c	13	111.8	2.6	901	29	BZ574671	BZ574671 msh2_3778
c	14	110.4	2.6	889	29	BZ551143	BZ551143 pacs1-60_
c	15	109.2	2.6	633	29	BZ551440	BZ551440 pacs1-60_
	16	107.8	2.5	827	29	BZ559631	BZ559631 pacs2-164
c	17	107.6	2.5	648	29	BZ548537	BZ548537 pacs1-60_
	18	107	2.5	709	28	BH488170	BH488170 BOGLI36TF
c	19	104	2.5	1429	29	BZ577585	BZ577585 msh2_5477
c	20	101.6	2.4	888	29	BZ567972	BZ567972 pacs2-164
	21	98.6	2.3	792	29	BZ574700	BZ574700 msh2_3790
	22	97	2.3	780	29	BZ573398	BZ573398 msh2_3086
	23	96.2	2.3	1342	29	BZ556627	BZ556627 pacs1-60_
	24	96	2.3	1122	29	BZ558365	BZ558365 pacs1-60_
	25	94.2	2.2	778	29	BZ573262	BZ573262 msh2_3024
c	26	93.4	2.2	915	29	BZ568304	BZ568304 pacs2-164
c	27	93.2	2.2	1307	29	BZ576975	BZ576975 msh2_5189
	28	92.8	2.2	628	12	BI180727	BI180727 LRH23D3 L
	29	92.2	2.2	580	29	BZ561162	BZ561162 pacs2-164
	30	91.6	2.2	885	29	BZ551313	BZ551313 pacs1-60_
c	31	91.6	2.2	1409	29	BZ577581	BZ577581 msh2_5475
	32	90.6	2.1	610	29	P974L	AL390681 Leishmani
	33	90.2	2.1	1388	29	BZ555697	BZ555697 pacs1-60_
	34	89.6	2.1	1118	29	BZ551441	BZ551441 pacs1-60_
	35	89.6	2.1	1150	29	BZ578915	BZ578915 msh2_6063
c	36	89	2.1	891	29	BZ567106	BZ567106 pacs2-164
c	37	88.8	2.1	850	29	BZ670377	BZ670377 PUBFG72TD
c	38	87.8	2.1	1131	29	BZ559111	BZ559111 pacs2-164
c	39	87	2.1	1489	29	BZ572227	BZ572227 msh2_2554
c	40	85.6	2.0	1154	29	BZ559289	BZ559289 pacs2-164
	41	84.8	2.0	1364	29	BZ567613	BZ567613 pacs2-164
	42	84.2	2.0	1594	29	BZ551753	BZ551753 pacs1-60_
	43	83.8	2.0	715	29	BZ560475	BZ560475 pacs2-164
	44	83.8	2.0	807	29	BZ568590	BZ568590 pacs2-164
	45	82.8	2.0	881	29	BZ549419	BZ549419 pacs1-60_

ALIGNMENTS

```
RESULT 1
BZ671349/c
LOCUS       BZ671349                 812 bp    DNA     linear   GSS 05-FEB-2003
DEFINITION  PUBKS59TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa077J21,
            genomic survey sequence.
ACCESSION   BZ671349
VERSION     BZ671349.1  GI:28219323
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 812)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
```

```
     source          1. .812
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa077J21"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
BASE COUNT      155 a    255 c    258 g    144 t
ORIGIN

  Query Match              4.1%;  Score 175.4;  DB 29;  Length 812;
  Best Local Similarity   52.7%;  Pred. No. 2.1e-26;
  Matches  433;  Conservative    0;  Mismatches  376;  Indels   12;  Gaps    2;

Qy       1651 TACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCA 1710
Db        812 TACGCCCAATTGAATAGCCGCGCCAACCGCCTAGCCCATCGCCTGATGGAAGCCGGCGTT 753

Qy       1711 CGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTT 1770
Db        752 GGGCCGGACGTACTGGTGGGGCTGGCCGTGGAGCGCTCCATCGACATGGTGGTCGGCCTG 693

Qy       1771 CTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAG 1830
Db        692 CTTGCCGTGCTCAAGGCCGGTGGCGCCTATGTGCCGCTGGACCCGGAATACCCGCGTGAG 633

Qy       1831 CGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTG 1890
Db        632 CGCTTGGCCTACATGCTGGAAGACAGTGGCGTGAAACTGCTGCTGACCCAGGCGCATCTG 573

Qy       1891 GATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTC 1950
Db        572 CGTGAACAGTTGCCGATTCCCCAAGGCCTGGAAACCCTGGTGTTGGGTGAGTCCGATTAC 513

Qy       1951 GAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTAT 2010
Db        512 GCCGGCTACAGCGACGCCAACCCAGGTATTCCC---CTGGATGGCGAAAACCTGGCCTAT 456

Qy       2011 GTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGT 2070
Db        455 GTGATCTATACCTCGGGCTCCACCGGCCAGCCTAAAGGCGCGGGCAACCGTCATTCGGCG 396

Qy       2071 GCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTG 2130
Db        395 TTGCTCAACCGCTTGCACTGGATGCAGGAAGCCTATGGCCTGGGCGCCAGCGACACGGTG 336

Qy       2131 CTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCG 2190
Db        335 CTGCAAAAGACCCCGTTCAGCTTCGACGTGTCGGTGTGGGAGTTTTTCTGGCCGCTGATG 276

Qy       2191 GCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCA 2250
Db        275 ACCGGCGCGCGCCTGGTGGTCGCTGCGCCGGGTGATCACCGTGATCCGGCCAAGCTGATC 216

Qy       2251 GAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATG 2310
Db        215 AACCTGATCAATCAACAACAGGTCACCACGCTGCACTTTGTGCCGTCGATGTTGCAGGCC 156

Qy       2311 CTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTG 2370
Db        155 TTC---------CTGCAGGATCCGGCAGTGTCGACCTGCCAGAGCTTGCAACGCATCGTG 105

Qy       2371 CTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGC 2430
Db        104 TGCAGCGGCGAAGCCTTGCCGGTGGATGCCCAGCAGCAGGTCTTCGCCAAGCTGCCGCAC 45

Qy       2431 GTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGAT 2471
Db         44 GCCGGGCTCTTCAACCTGTACGGCCCGACCGAAGCGGCCAT 4
```

```
RESULT 2
BZ563956
LOCUS       BZ563956                 627 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_4474.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_4474, genomic survey sequence.
ACCESSION   BZ563956
VERSION     BZ563956.1  GI:27188838
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_4474"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      103 a    220 c    199 g    104 t      1 others
ORIGIN

  Query Match              4.1%;  Score 174.4;  DB 29;  Length 627;
  Best Local Similarity   60.9%;  Pred. No. 3.2e-26;
  Matches  300;  Conservative    0;  Mismatches  192;  Indels    1;  Gaps    1;

Qy       1997 CGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGA 2056
Db        117 CCGAGCGCGGCTCCGTGAGCTACACGTCCGGCTCCACCGGCACGCCAAAGGGCGTGATGC 176

Qy       2057 TCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGC 2116
Db        177 TCAGCCATGCGGGGGTGAGCAACACGCTGCTCGACATCAACCAGCGCTACGGCGTCGACG 236

Qy       2117 CCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGT 2176
Db        237 CCAACGACCGCGTCCTCGGCCTCGCCGAGCTGAGCTTCGACCTCTCGGTCTACGACTTCT 296

Qy       2177 TCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATC 2236
Db        297 TCGGGGCCACCGCGGCGGTGGCCCAGGTGGACCTTCGCGACGCTGCGCGCGAGAACGATC 356

Qy       2237 CGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGG 2296
Db        357 CATCTGACGGGCCTGTGCTGCTGGAACGCCACGCCATCACCCTGCGGAACTCGGTGCCGG 416

Qy       2297 CGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTC 2356
Db        417 CCCAAGGCCAGATGCTCATCGATTACCTGGAGAGCGAGCCGCAACGTCACCTGCCGGGAC 476

Qy       2357 TGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGG 2416
Db        477 CGAGCTGCGTGCTCTGGTCCGGNGACTGGATTCCGGTCAGCCTGCC-GACCCGCTGTGGC 535

Qy       2417 CCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGT 2476
```

```
Db        536 GGTGCTGGCCGGACAGCTCGCCTGTTAGCCTGGGCGGAGCCACCGAGGCGGCGATCTGGC 595

Qy       2477 CCATCGGGTACCC 2489

Db        596 CGATCGAACATCC 608
```

```
RESULT 3
BZ572480/c
LOCUS       BZ572480                1552 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  msh2_2654.x1 msh Pseudomonas aeruginosa genomic clone msh2_2654,
            genomic survey sequence.
ACCESSION   BZ572480
VERSION     BZ572480.1  GI:27207541
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1552)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..1552
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_2654"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."
BASE COUNT      320 a    448 c    387 g    390 t      7 others
ORIGIN

  Query Match            3.7%;  Score 157.6;  DB 29;  Length 1552;
  Best Local Similarity  54.4%;  Pred. No. 1.3e-22;
  Matches  356;  Conservative    0;  Mismatches  275;  Indels   24;  Gaps    1;

Qy       1653 CGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACG 1712

Db        724 CGACGAACTGTCCGAGCGTGCACTGCGCATCGCCGGCGGCTTGCACGAAGCCGGGGTGCG 665

Qy       1713 CCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCT 1772

Db        664 CCCTGGCGACGCGGTCGAGGTCAGCCTGCCGCGCGGACCGCAGCAGGTCGCGNCGGTATT 605

Qy       1773 CGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCG 1832

Db        604 CGGCGTGCTCGCCGCAGGCGCCTGCTACGTGCCGCTGGACATCGACCAGCCGCCCGCACG 545

Qy       1833 TATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGA 1892

Db        544 GCGGCGCCTGATCGAAGAGGCCGCCGGGGTATGCCTGGCGATCACCGAGGAGGACGATCC 485

Qy       1893 TGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGA 1952

Db        484 GCAGGCCTTGCCGCCGCGCCTGGATGTCCAGCGCCTGCTGCGCGGCCCGGCGCTGGCCGC 425

Qy       1953 AGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGT 2012

Db        424 CCCCGTGCCG------------------------CTGGCGCCGCAGGCGAGTGCCTATGT 389

Qy       2013 CATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGC 2072

Db        388 GATCTACACCTCGGGCTCCACCGGGGTGCCCAAGGGCGTCGAGGTCAGCCACGCGGCGGC 329

Qy       2073 CGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCT 2132

Db        328 GATCAATACCATCGACGCGCTGCTCGACCTGCTGCGGGTGAACGCATCGGATCGCTTGCT 269

Qy       2133 GGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGC 2192

Db        268 GGCGGTCTCCGCGCTGGACTTCGATCTGTCGGTCTTCGACCTGTTCGGCGGCCTCGGCGC 209

Qy       2193 GGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGA 2252

Db        208 CGGTGCCAGCCTGGTCCTGCCGGCCCAGGAACAGGCGCGCGATGCCGCTGCCTGGGCGGA 149

Qy       2253 GTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGG 2307

Db        148 GGCTATCCAGCGGCATGCGGTGAGCCTGTGGAACTCGGCGCCGGCCTTGCTGGGG 94
```

```
RESULT 4
BZ562842/c
LOCUS       BZ562842                765 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_3957.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_3957, genomic survey sequence.
ACCESSION   BZ562842
VERSION     BZ562842.1  GI:27185526
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 765)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..765
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_3957"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      157 a    234 c    248 g    125 t      1 others
ORIGIN

  Query Match            3.1%;  Score 133;  DB 29;  Length 765;
  Best Local Similarity  52.0%;  Pred. No. 1.5e-17;
  Matches  323;  Conservative    0;  Mismatches  295;  Indels    3;  Gaps    1;

Qy       1691 GGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCT 1750

Db        673 GTCTGCGCGAGGAGGGTGTCGGCATCAATGTACTGGTCGGTATCGCGCTGGAGCACGGCG 614

Qy       1751 GGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCG 1810

Db        613 TACCCATGGTGGTGGCCCTGCTGACGGTGCTCAAGACCGGTGGCGCCTATGTGCCGCTGG 554
```

```
Qy       1811 ATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCG 1870

Db        553 ATCCGCAGTATCCCGCGGACCGCTTGCAGTACATGATCGACGACAGCGGACTGCGCCTGT 494

Qy       1871 TGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGC 1930

Db        493 TGCTCAGCCAGCAAAGCGTGCTGGCGCGCCTGCCGCAGTCGGATGGGCTGCAAAGCCTGC 434

Qy       1931 TCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTC--- 1987

Db        433 TGCTCGACGACCTGGAGCGACTGGTGCACGGCTACCCGGCGGAAAACCCGGACCTGCCGG 374

Qy       1988 AGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGG 2047

Db        373 AGGCGCCCGACAGTCTCTGCTACGCGATCTACACCTCCGGTTCCACCGGCCAGCCGAAAG 314

Qy       2048 GGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCG 2107

Db        313 GCGTGATGGTCCGCCATCGAGCCTTGACCAACTTCGTCTGCAGCATCGCCCGGCAACCGG 254

Qy       2108 AAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCT 2167

Db        253 GCATGCTGGCGCGGGACCGTCTGCTGTCGGTGACCACCTTCTCCTTCGATATCTTCGGCC 194

Qy       2168 ATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGC 2227

Db        193 TGGAGCTATACGTACCGCTGGCGCGTGGCGCCAGCATGCTGCTCGCCAGCCGCGAGCAGG 134

Qy       2228 TGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACT 2287

Db        133 CCCAGGATCCCGAGGCCCTGCTCGATCTGGTCGAGCGAACGGGGGATCCACTAGTTCTAG 74

Qy       2288 CGGTGCCGGCGCTGATGCGGA 2308

Db         73 AGCGGCCGCCACCGCGGTGGA 53
```

```
RESULT 5
AQ989512/c
LOCUS       AQ989512                646 bp    DNA     linear   GSS 14-AUG-2000
DEFINITION  Rfc00062 Photorhabdus luminescens strain W14 M13 library
            Photorhabdus luminescens genomic clone PLG00062, genomic survey
            sequence.
ACCESSION   AQ989512
VERSION     AQ989512.1  GI:9648106
KEYWORDS    GSS.
SOURCE      Photorhabdus luminescens
  ORGANISM  Photorhabdus luminescens
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Photorhabdus.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   ffrench-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T.,
            Daborn,P.J., Bowen,D. and Blattner,F.R.
  TITLE     A genomic sample sequence of the entomopathogenic bacterium
            Photorhabdus luminescens W14: potential implications for virulence
  JOURNAL   Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
  MEDLINE   20378633
   PUBMED   10919786
COMMENT     Contact: ffrench-Constant RH
            Department of Biology and Biochemistry
            University of Bath
            South Building, Bath BA2 7AY, UK
            Tel: (44) 1225 826621
            Fax: (44) 1225 826779
            Email: bssrfc@bath.ac.uk
            This is one of 2,122 random reads from the M13 library. For
            annotation of identified clones (BLASTX, BLASTN and mapping to E.
            coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic
            Acids Res.
            Seq primer: M13 Forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..646
                     /organism="Photorhabdus luminescens"
                     /mol_type="genomic DNA"
                     /strain="W14"
                     /db_xref="taxon:29488"
                     /clone="PLG00062"
                     /dev_stage="primary phase variant"
                     /clone_lib="Photorhabdus luminescens strain W14 M13
                     library"
                     /note="Genomic DNA from strain W14 was size selected (1-2
                     kb) and then cloned into M13 Janus."
BASE COUNT      177 a     163 c     151 g     150 t       5 others
ORIGIN

  Query Match               3.0%;  Score 128.8;  DB 28;  Length 646;
  Best Local Similarity   54.7%;  Pred. No. 1.1e-16;
  Matches  256;  Conservative    0;  Mismatches  212;  Indels    0;  Gaps    0;

Qy       1973 CGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCA 2032

Db        469 CAAATATGGCTAAGCGTGCGGTGACTGATCCGGCTTATATTATCTATACCTCTGGCTCAA 410

Qy       2033 CAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACA 2092

Db        409 CCGGTATGCCAAAAGGGGTGGTTATCAGCCACCAAAGCGCACTCAATACCTGTCTGGATA 350

Qy       2093 TCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCT 2152

Db        349 TTAATCGCCGCCATCAGGTTCAAAGTAGCGACCGATTATTGGCACTCTCCGCCCTACATT 290

Qy       2153 TCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGC 2212

Db        289 TCGACCTCTCCGTTTATGACATTTTCGGTATTCTCACGGCGGGAGGCGCGTTGGTATTAA 230

Qy       2213 CGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGG 2272

Db        229 TTAATGAAGCACAGCGTCGTGATCCATCAACCTGGGAAATGTTGATCTCTCGTTACAACG 170

Qy       2273 TGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTC 2332

Db        169 TAACGATATGGAACAGCGTACCGGCACTGTTCGACATGTTACTGACCTATTGTGAAGGGA 110

Qy       2333 GCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGG 2392

Db        109 TGGAAGTGGATGCACCGAAAGTACTTCGTACCGTGATGTTGTCCGGGGACTGGATTGGTT 50

Qy       2393 TGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGA 2440

Db         49 TATCACTGCCTGAACGTTACCGGGCATTCAATCCTGATGGTGTGTTGA 2
```

```
RESULT 6
BZ548527
LOCUS       BZ548527               1117 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs1-60_1192.s1 pacs1-60 Pseudomonas aeruginosa genomic clone
            pacs1-60_1192, genomic survey sequence.
ACCESSION   BZ548527
VERSION     BZ548527.1  GI:27152108
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1117)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol., (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
```

```
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..1117
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="1-60"
                     /db_xref="taxon:287"
                     /clone="pacs1-60_1192"
                     /clone_lib="pacs1-60"
                     /note="clinical isolate 1-60 Whole genomic shotgun
                     library."
BASE COUNT      169 a    403 c    330 g    212 t      3 others
ORIGIN

  Query Match             3.0%;  Score 125;  DB 29;  Length 1117;
  Best Local Similarity   52.5%;  Pred. No. 7.9e-16;
  Matches  323;  Conservative    0;  Mismatches  285;  Indels    7;  Gaps    2;

Qy        845 AGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTGAAGC 904
Db         53 AGCCGCGCACCCGGCGCCTGGCATTCCAGCTTTCCGCCGGCGAGAGCCGGCTACTCGAGC 112

Qy        905 GGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCCGAGG 964
Db        113 GTCTTGCCGCGCACAGTGGCGTGACCTTGTCCAGCGTGTTCGGCTGCGCCTTCGCGCTGG 172

Qy        965 TGATCGGGCGCTGGAGCGCGAGCCCC-CGGTTTACGCTCAACATAACGCTCTTCAACCGG 1023
Db        173 TCCTGGCGCGCTGGAGCGAAAGCGCGGCAATTTCTCCTCAACGTGCCGTTGTTCGATCGG 232

Qy       1024 CTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTCCTG 1083
Db        233 CATGCCGACGACCCGCGTATCGGCGAGGTGATCGCCGACTTCACCACCCTGTTGCTGCTG 292

Qy       1084 GACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAGCAG 1143
Db        293 GAGTGCCGGATGCAGGCCGGGGTGTCCTTCGCCGAGGCGGTGAAGAGCTTCCAGCGCAAC 352

Qy       1144 CTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCCGCC 1203
Db        353 CTCCACGGAGCCATCGACCACGCCGCATTCCCCGCCCTGGAGGTGCTCCGCGAGGC---- 408

Qy       1204 CGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTTAAC 1263
Db        409 --GCGCCGGCAGGGCCAGCCACGCTCGGCGCCGGTGGTGTTCGCCAGCAACCTGGGCGAG 466

Qy       1264 CAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGCACG 1323
Db        467 GAGGGCTTCGTCCCGGCGGCCTTCCGCGACGCTTTCGGCGATCTCCACGACATGCTCTCG 526

Qy       1324 CAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTCCTC 1383
Db        527 CAGACCCCGCAAGTCTGGCTCGACCACCAGCTCTACCGGGTGGGCGACGGTATCCTGCTG 586

Qy       1384 GCGTGGGACATCGTCGACGGAGTGTTCCCCGCCCGACCTTCTGGACGACATGCTCGAAGCG 1443
Db        587 GCCTGGGACAGCGTCGTCGGCCTGTTCCCCGAGAGCCTGCCGGAAACCATGTTCGAAGCC 646

Qy       1444 TACGTCGTTTTTCTC 1458
Db        647 TACGTGGGGCTGCTC 661


RESULT 7
BZ570019
LOCUS       BZ570019                 921 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  msh2_1128.y2 msh Pseudomonas aeruginosa genomic clone msh2_1128,
            genomic survey sequence.
ACCESSION   BZ570019
VERSION     BZ570019.1  GI:27205080
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol., (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..921
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_1128"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."
BASE COUNT      175 a    264 c    268 g    201 t     13 others
ORIGIN

  Query Match             2.9%;  Score 121.6;  DB 29;  Length 921;
  Best Local Similarity   52.3%;  Pred. No. 3.9e-15;
  Matches  293;  Conservative    0;  Mismatches  264;  Indels    3;  Gaps    1;

Qy       1574 TGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGG 1633
Db        145 TGCACCGGTTGTTCGAGGAGCAGGTCGAGCGCACGCCGACGGCGCCGGCGCTGGCCTTCG 204

Qy       1634 CGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGC 1693
Db        205 GCGAGGAGCGCCTGGACTACGTCGAGCTGAACCGCCGTGCCAACCGCCTGGCGCATGCCC 264

Qy       1694 TGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGG 1753
Db        265 TGATCGAGCGCGGGATCGGCGCGGACCGCCTGGTGGGCGTGGCCATGGAACGTTCCATCG 324

Qy       1754 AGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATG 1813
Db        325 AGATGGTCGTGGCCCTGATGGCGATCCTCAAGGCCGGCGGCGCCTACGTGCCGGTGGACC 384

Qy       1814 CCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGC 1873
Db        385 CGGAGTATCCCGAGGAGCGCCAGGCCTACATGCTGGAGGACAGCGGCGTGCAACTGCTGC 444

Qy       1874 TGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCG 1933
Db        445 TCAGCCAGTCGCACCTGAAGCTGCCGCTGGCGCAAGGTGTGCAGCGGATCGACCTGGACC 504

Qy       1934 TGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACAC 1993
Db        505 AGGCCGATGCCTGGCTGGAAAACCATGCCGAGAACAATCCAGGGGT---CGAGCTGAACG 561

Qy       1994 CTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGA 2053
Db        562 GCGAGAATCTTGCCTATGTCATCTACACCTCCGGTTCCACCGGCAAGCCCAAAGGCGCCG 621

Qy       2054 TGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAG 2113
Db        622 GCAACCGCCATTCGGCGCTGAGCAACCGGTTGTGCTGGATGCAGCAGGCCTATGGGCTGG 681

Qy       2114 GGCCCGGAGACAGGGTGCTG 2133
```

```
Db        682 GCGGTTGCGACACGGTGCTG 701

RESULT 8
BZ567332
LOCUS       BZ567332                1038 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_6886.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_6886, genomic survey sequence.
ACCESSION   BZ567332
VERSION     BZ567332.1  GI:27198451
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1038)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..1038
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_6886"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      187 a    330 c    304 g    216 t      1 others
ORIGIN

  Query Match             2.9%;  Score 120.8;  DB 29;  Length 1038;
  Best Local Similarity   50.0%;  Pred. No. 5.9e-15;
  Matches  328;  Conservative    0;  Mismatches  327;  Indels    1;  Gaps    1;

Qy       1558 CTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAG 1617
Db         80 CCGGACGAGCGCAGCGTCGAGCAGCTGTTCGCCAGCCGCGCGGCGATGGAGCCCGAGCGG 139

Qy       1618 CTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGG 1677
Db        140 GTGGCATTGCTCTTCGAGGAGCGCCAGCTGAGCTACGGCGAGCTCAACGCCCAGGCCAAC 199

Qy       1678 CGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTG 1737
Db        200 CGCCTGGCCCACCGCCTGATCGAGCTGGGCGTGGGGCCTGACGTGCTGGTTGGGATCGCC 259

Qy       1738 ATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCC 1797
Db        260 GTCGAGCGCGGGCTGGAAATGATCGTCAGCCTGCTGGCGGTGCTCAAGGCCGGCGGCGCC 319

Qy       1798 TACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGT 1857
Db        320 TACGTACCGCTGGACCCGGAATATCCGCAGGAGCGCCTCGGCTACATGATCGAGGACAGC 379

Qy       1858 GAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGG 1917
Db        380 GGCATCGCCCTGTTGCTCAGCCAGAGCCACCTGCTGCAACGCCTGCCAGCGGCCTCTGGG 439

Qy       1918 ATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATG 1977
Db        440 ATCGCCTGCCTGGCGCTCGACCAGGCGTGGGACTGGCAGGACCGTCCGGCGAGCGACCCG 499

Qy       1978 ATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGG 2037
Db        500 CAGTTGCGGGCGCATCCGCAGAACCTCGCCTACGTGATGTTCACCTCCGGTTCCACCGGC 559

Qy       2038 TTGCCCAAGGGGGT-GATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAA 2096
Db        560 CGTCCCAAGGGCGTCGGCAATAGCCGAGAGAGCCTGTCGCGGCATACCCACGTTTCGCTT 619

Qy       2097 CGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGA 2156
Db        620 GGAGTTATTTGGCATCGGGCCGGACGACCGGGTGCTGCAATTCTCCACCTTCAATTTCGA 679

Qy       2157 TCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGC 2212
Db        680 CGGGTTCGTCGAGCAGCTCTACCCACCGTTCGCCTTGGGTGCCTCGGTGGTGCTGC 735

RESULT 9
BZ548604
LOCUS       BZ548604                 667 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs1-60_1266.s1 pacs1-60 Pseudomonas aeruginosa genomic clone
            pacs1-60_1266, genomic survey sequence.
ACCESSION   BZ548604
VERSION     BZ548604.1  GI:27152185
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 667)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..667
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="1-60"
                     /db_xref="taxon:287"
                     /clone="pacs1-60_1266"
                     /clone_lib="pacs1-60"
                     /note="clinical isolate 1-60 Whole genomic shotgun
                     library."
BASE COUNT      100 a    235 c    234 g     98 t
ORIGIN

  Query Match             2.8%;  Score 119;  DB 29;  Length 667;
  Best Local Similarity   52.1%;  Pred. No. 1.3e-14;
  Matches  266;  Conservative    0;  Mismatches  245;  Indels    0;  Gaps    0;

Qy       2438 TGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGA 2497
Db        133 TGGTCAATTTCTACGGGCCGACCGAGGACACCACCTATTCGACGGCGCACGAACTGGACC 192

Qy       2498 ACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCC 2557
Db        193 TGCACGCCGAAGCGCTCGACGAGCCGCCCATCGGCCGACCGCTGCCCGGCACCACGGTGG 252

Qy       2558 ACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTG 2617
Db        253 AGGTGCTCGACGGCTTCGAGGCGCCGCTGCCGCTGGGCGTGGCCGGGGAGCTGTACCTCG 312
```

```
Qy       2618 GCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCC 2677
Db        313 GCGGCATCGGCCTGGCGCGTGGCTATTTCGGCAAGCCGGAGCAGACCGCCGAGCGTTTCC 372

Qy       2678 TCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGC 2737
Db        373 GCGTCGATCCGGGCAGCGGCGAGCGCCGCTATCGCACCGGCGACCGCGTACGCATGCGCG 432

Qy       2738 CCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACC 2797
Db        433 AGGACGGCGTGCTGGAACACCTCGGCCGGCTCGACGACCAGGTCAAGTTCAACGGCTTCC 492

Qy       2798 GCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGG 2857
Db        493 GCATCGAGCTGGGTGAGATCGCCTCGTGCCTGGCGAGTTTTCCCGGGGTGAGCGAACGCT 552

Qy       2858 TGATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGG 2917
Db        553 GCGCCATGCTCACCGAGGACAGCGCCGGCCTGCGTCGGCTGGTCGGCTACCTGGCGGCGC 612

Qy       2918 AGGGCACACGGAGACGCGCTGCCGAGCAGGA 2948
Db        613 CGTTCGCACCGCCGCTCCAGGCGCTGAACGA 643
```

```
RESULT 10
BZ561650/c
LOCUS       BZ561650                 753 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_3367.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_3367, genomic survey sequence.
ACCESSION   BZ561650
VERSION     BZ561650.1  GI:27182090
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 753)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .753
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_3367"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      129 a    257 c    244 g    123 t
ORIGIN

  Query Match              2.8%;  Score 117;  DB 29;  Length 753;
  Best Local Similarity   52.9%;  Pred. No. 3.4e-14;
  Matches  353;  Conservative    0;  Mismatches  295;  Indels   19;  Gaps    4;

Qy       2195 GCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGT 2254
Db        732 GGGGCACGTTGGTGGTGACCGCCCCGGCGGCCATTCCCGACCCGGCGAAGCTGGTGCCCC 673

Qy       2255 TGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCG 2314
Db        672 TGATCACCCGCAAAGGGGTCGATACGCTGCACTTCGTGCGGTCAATGCTGCAGGCCTTCT 613

Qy       2315 TCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGA 2374
Db        612 TGCAGGATGAAGACGTCGCCTCCTGCACCAGC---------TTGAAACGCATCGTTTGCA 562

Qy       2375 GCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGT 2434
Db        561 GCGGCGAGGCGCTGCCGGCGAACGCCCAGCAGCAGGTGTTCGCCAAGCTGCCGCAGGCCG 502

Qy       2435 CGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGA 2494
Db        501 GCCTCTATAACCTCTATGGCCCGACCGAGGCGGCCAT------CGACGTCACCCACTGGA 448

Qy       2495 GGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGT 2554
Db        447 CCTGCGTGGAGGAGGGCAAGGACGCGGTGCCGATCGGCCGGCCGATCGCCAACCTGGCCT 388

Qy       2555 TCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACA 2614
Db        387 GCTACATCCTCGATGGCAACCTGGAGCCGGTGCCGGTGGGCGTGCTCGGCGAGCTGTACC 328

Qy       2615 TTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCT 2674
Db        327 TGGCCGGTCGGGGCCTGGCCCGTGGCTACCACCAGCGTCCGGGGCTGACTGCCGAGCGTT 268

Qy       2675 TCCTCGTGCACCC---CGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCT 2731
Db        267 TCGTCGCCAGCCCGTTCGTGGCTGGGGAGCGGATGTACCGCACCGGCGATTTGGCGCGCT 208

Qy       2732 ACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCG 2791
Db        207 ACCGCGCCGACGGGGTGATCGAGTACGCCGGGCGGATCGACCACCAGGTGAAGCTGCGCG 148

Qy       2792 GATACCGCGTTGAGCTCGGGGAAATCGAGG-AAACGCTCAAGTCGCATCCGAACGTACGC 2850
Db        147 GCTTGCGCATCGAGCTGGGCGAGATCGAGGCAAAGCCTGCTGGAGCATCCGTGGGTGCGC 88

Qy       2851 GACGCGG 2857
Db         87 GAGGGGG 81
```

```
RESULT 11
BZ558891
LOCUS       BZ558891                 840 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_109.s1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_109, genomic survey sequence.
ACCESSION   BZ558891
VERSION     BZ558891.1  GI:27174215
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 840)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Pseudomonas aeruginosa"
```

```
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_109"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      113 a    296 c    272 g    159 t
ORIGIN

  Query Match             2.7%;  Score 114.4;  DB 29;  Length 840;
  Best Local Similarity   53.1%;  Pred. No. 1.2e-13;
  Matches  292;  Conservative    0;  Mismatches  251;  Indels    7;  Gaps    2;

Qy        935 CGGGCGTCATCCTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGT 994
Db         45 CGAGCGTGTTCGGCTGCGCCTTCGCGCTGGTCCTGGCGCGCTGGAGCG-AAACGCGGAAT 103

Qy        995 TTACGCTCAACATAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCA 1054
Db        104 TTCTCCTCAACGTGCCGTTGTTCGAGCGGCATGCCGACGACCCGCGTATCGGCGAGGTGA 163

Qy       1055 CCGGGGACTTCACGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCG 1114
Db        164 TCGCCGACTTCACCACCCTGTTGCTGCTGGAGTGCCGGATGCAGGCCGGGGTGTCCTTCG 223

Qy       1115 AACAGCGCGCTAAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAA 1174
Db        224 CCGAGGCGGTGAAGAGCTTCCAGCGCAACCTCCACGGAGCCATCGACCACGCCGCATTCC 283

Qy       1175 GCGGTATCGAGGTCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGT 1234
Db        284 CCGCCCTGGAGGTGCTCCGCGAGGC------GCGCCGGCAGGGCCAGCCACGCTCGGCGC 337

Qy       1235 TCCCCGTGGTGCTCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGA 1294
Db        338 CGGTGGTGTTCGCCAGCAACCTGGGCGAGGAGGGCTTCGTCCCGGCGGCCTTCCGCGACG 397

Qy       1295 GGCTCGGAACTCCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGC 1354
Db        398 CTTTCGGCGATCTCCACGACATGCTCTCGCAGACCCCGCAGGTCTGGCTCGACCACCAGC 457

Qy       1355 TCTACGAGCACGATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGC 1414
Db        458 TCTACCGGGTGGGCGACGGTATCCTGCTGGCCTGGGACAGCGTCGTCGGCCTGTTCCCCG 517

Qy       1415 CCGACCTTCTGGACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGG 1474
Db        518 AAGGTCTGCCGGAAACCATGTTCGAAGCCTACGTGGGGCTGCTCCAGCGTCTCTGCGACA 577

Qy       1475 AACCATGGGG 1484
Db        578 CGCCTGGGAG 587


RESULT 12
BZ680689
LOCUS       BZ680689                 827 bp    DNA     linear   GSS 05-FEB-2003
DEFINITION  PUBEL33TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa036F18,
            genomic survey sequence.
ACCESSION   BZ680689
VERSION     BZ680689.1  GI:28235748
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 827)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..827
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa036F18"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
BASE COUNT      147 a    294 c    241 g    145 t
ORIGIN

  Query Match             2.7%;  Score 114;  DB 29;  Length 827;
  Best Local Similarity   50.8%;  Pred. No. 1.5e-13;
  Matches  301;  Conservative    0;  Mismatches  285;  Indels    6;  Gaps    1;

Qy        569 ACGAGCGGCAAACCCGTCTCGTGCTCAGTATCGATCTCATTAACGTTGACCTAGGCAGCC 628
Db        241 ACGGCCGGCACACACGCCTGGCGATCAGCCTGGACAATTTGATCCTCGACGCCCTGAGCA 300

Qy        629 TGTCCATCATCTTCAAGGACTGGCTCAGCTTCTACGAAGATCCCGAGACCTCTCTCCCTG 688
Db        301 TCCTGCGTTTCTACGCCGAACTCGATGCGTTGTACCGCGAGCCCGACATAGCGCTGGCCC 360

Qy        689 TCCTGGAGCTCTCGTACCGCGATTATGTACTCGCGCTGGAGTCTCGCAAGAAGTCTGAGG 748
Db        361 CGTTGGGTCTGTCGTTTCGCGATTA------CCAATTGCAGAGCGCCCCGGCACCTGAAG 414

Qy        749 CGCATCAACGATCGATGGATTACTGGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGA 808
Db        415 AGCTGCACGCCGCGCAGCGCTTCTGGCAGGAACGCCTGCCCGAATTGCCGCCCCATCCGC 474

Qy        809 CGCTTCCGATGAAGGCCGATCCATCTACCCTGAAGGAGATCCGCTTCCGGCACACGGAGC 868
Db        475 AGTTGCCGCTGGCCGCCGACCCCGCCAGCCTGGGCCGACCACGCTTTGAACGCCTGCAAG 534

Qy        869 AATGGCTGCCGTCGGACTCCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGA 928
Db        535 GGCAGATCGATGCCAGCGCCTGGCAAGCGATCCTGTCCAAGGCCCAGCAACAGGGGCTCA 594

Qy        929 CCCCGACGGGCGTCATCCTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCC 988
Db        595 CCGCCTCGGCGGTATTGCTCTGCGCTTTTGCCGAAACCCTCGGCCGCTGGAGTGCACGCC 654

Qy        989 CCCGGTTTACGCTCAACATAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACG 1048
Db        655 CGGGCCTGAGCCTAAACCTCACACTGTTCGATCGCCGCCCCGTACACCCTCAGATCGACC 714

Qy       1049 ATATCACCGGGGACTTCACGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGA 1108
Db        715 AGGTCATGGGCGACTTCACATCGCTGACCCTGCTCGGGTATGTGCCCGTGGCGGGTGAGC 774

Qy       1109 GCTTCGAACAGCGCGCTAAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGA 1160
Db        775 GCTGGGTTGACCGTGCACGACGTACGCAACAATCCCTGGGCGAGGCGTTGGA 826


RESULT 13
BZ574671/c
LOCUS       BZ574671                 901 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  msh2_3778.x1 msh Pseudomonas aeruginosa genomic clone msh2_3778,
            genomic survey sequence.
ACCESSION   BZ574671
VERSION     BZ574671.1  GI:27209732
```

```
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 901)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .901
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_3778"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."
BASE COUNT      140 a    305 c    314 g    140 t      2 others
ORIGIN

  Query Match             2.6%;  Score 111.8;  DB 29;  Length 901;
  Best Local Similarity   52.4%;  Pred. No. 4.3e-13;
  Matches  273;  Conservative    0;  Mismatches  242;  Indels    6;  Gaps    1;

Qy       1536 CGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCG 1595
Db        601 CGCCGGCACGGCCGGCGAGGCAGGGCTCCAGGACACCCTCCATGGCCTGTTCGCCGCCCG 542

Qy       1596 GGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGA 1655
Db        541 CGTGGCGGCCAGCCCACAGGCCCCGGCGCTGACCTTCGCCGGACAGACCCTGAGCTATGC 482

Qy       1656 AGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCC 1715
Db        481 CGAACTCGATGCCCGCTCCAACCGTCTGGCGCGCGTCCTGCGCAGTCACGGCGTCGGCCC 422

Qy       1716 GAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGC 1775
Db        421 GGAAGTGCGGGTCGGCCTGGCCCTGGAGCGTTCGCTGGAGATGGTCGTCGGCCTGCTGGC 362

Qy       1776 GGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTAT 1835
Db        361 GATCCTCAAGGCCGGCGGCGCCTACGTGCCGCTGGACCCGGAATATCCGCTGGAGCGCCT 302

Qy       1836 CCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGG 1895
Db        301 GCAATACATGATCGAGGACAGCGGCGTACGCCTGCTGCTCAGCCACGCGGCGCTGTTCGA 242

Qy       1896 CAAACTGTCATGGCCGCCGGGGATCCAGCGGCTG------CTCGTGAGCGAGGCCGGCGT 1949
Db        241 GGCCCTCGGCGAGCTGCCGGCAGGCGTGGCGCGCTGGTGCCTCGAGGAGGACGGCCCGGC 182

Qy       1950 CGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTA 2009
Db        181 GCTGGACGCCGAGGACCCCGCGCCGCTGGCAGCGCTGAGCGGGCCGCAGCACCAGGCCTA 122

Qy       2010 TGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGG 2050
Db        121 CCTGATCTATACCTCCGGCTCCACCGGCAAGCCGAAGGGGG 81
```

```
RESULT 14
BZ551143/c
LOCUS       BZ551143               889 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs1-60_3008.x1 pacs1-60 Pseudomonas aeruginosa genomic clone
            pacs1-60_3008, genomic survey sequence.
ACCESSION   BZ551143
VERSION     BZ551143.1  GI:27154724
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .889
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="1-60"
                     /db_xref="taxon:287"
                     /clone="pacs1-60_3008"
                     /clone_lib="pacs1-60"
                     /note="clinical isolate 1-60 Whole genomic shotgun
                     library."
BASE COUNT      164 a    294 c    280 g    148 t      3 others
ORIGIN

  Query Match             2.6%;  Score 110.4;  DB 29;  Length 889;
  Best Local Similarity   50.8%;  Pred. No. 8.5e-13;
  Matches  349;  Conservative    0;  Mismatches  322;  Indels   16;  Gaps    3;

Qy       2002 CTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGAT 2061
Db        800 CTGGCTTATGTCTTCTCCACTCGCGGTCCCACCGTTAAGCCCAAGGGGGCCGCCAACCGC 741

Qy       2062 CATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGA 2121
Db        740 CATTGGGCGTTAAGCAACGGCTTGTGCTGGATGCAGCAGNCTTATGGGCTGGGCGTCGGC 681

Qy       2122 GACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGG 2181
Db        680 ACCACGGTGTTGCAGAAGACCCCGTTCAGCTTCGACGTGTCGGTCTGGGAGTTCTTCTGG 621

Qy       2182 ATCCTGGCGGCGGGCGGTACG-ATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGC 2240
Db        620 GCCGCTGATGAGCGGGGCGCGTTTGGTGGTGACCGCGCCGGGCGACCATCGCGATCCGGC 561

Qy       2241 GCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCT 2300
Db        560 GAAGCTGGTGGCGCTGATCAATCGCGAAGGGGTCGACACGCTGCACTTCGTGCCGTCGAT 501

Qy       2301 GATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCG 2360
Db        500 GCTGCAGGCCTTCCTGCAGGACGAAGACGTCGCCTCCTGCACCAGC---------CTGAA 450

Qy       2361 GCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCAT 2420
Db        449 ACGCATCGTTTGCAGCGGCGAGGCGCTGCCGGCGGACGCCCAGCAGCAGGTGTTCGCCAA 390

Qy       2421 CAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCAT 2480
```

```
Db         389 GCTGCCGCAGGCCGGCCTCTATAACCTCTATGGCCCGACCGAGGCGGCCATCGACGTCAC 330

Qy        2481 CGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCT 2540

Db         329 CCACTGGACCTGCATGGAGGAGGGCAAGGACGCGGTGCCGAT------CGGCCGGCCGAT 276

Qy        2541 GCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCC 2600

Db         275 CGCCAACCTGGCCTGCTACATCCTCGATGGCAACCTGGAGCCGGTACCGGTGGGCGTGCT 216

Qy        2601 GGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAA 2660

Db         215 CGGCGAGCTGTACCTGGCCGGTCGGGGCCTGGCCCGTGGCTACCACCAGCGTCCGGGGCT 156

Qy        2661 GACGCGCAAGAGCTTCCTCGTGCACCC 2687

Db         155 GACTGCCGAGCGCTTCGTCGCCAGCCC 129


RESULT 15
BZ551440/c
LOCUS       BZ551440                 633 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs1-60_3153.x1 pacs1-60 Pseudomonas aeruginosa genomic clone
            pacs1-60_3153, genomic survey sequence.
ACCESSION   BZ551440
VERSION     BZ551440.1  GI:27155021
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 633)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..633
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="1-60"
                     /db_xref="taxon:287"
                     /clone="pacs1-60_3153"
                     /clone_lib="pacs1-60"
                     /note="clinical isolate 1-60 Whole genomic shotgun
                     library."
BASE COUNT       97 a    226 c    214 g     96 t
ORIGIN

  Query Match             2.6%;  Score 109.2;  DB 29;  Length 633;
  Best Local Similarity   56.1%;  Pred. No. 1.4e-12;
  Matches  229;  Conservative    0;  Mismatches  173;  Indels    6;  Gaps    1;

Qy        2523 CCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACC 2582

Db         554 CCGATCGGGCCGGCCGCTGGGCAACGTGGTCTGCCGCGTGCTGGACGCCGAGTTCAACCT 495

Qy        2583 GCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTA 2642

Db         494 GCTGCCGGCTGGCGTCGCCGGCGAGCTGTGCATCGGCGGCCTGGGCCTGGCCCGCGGCTA 435

Qy        2643 CTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCG------TGCACCCCGAGACCGG 2696

Db         434 CCTGGGGCGCCCGGCGCTGAGCGCCGAGCGCTTCGTCGCCGATCCGTTCTCGGCCGCTGG 375

Qy        2697 GGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTT 2756

Db         374 CGAGCGCCTGTACCGCACCGGCGACCGCGCGCTGGAACGCCGACGGCGTGCTGGAATA 315

Qy        2757 CATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAAT 2816

Db         314 CCTCGGCCGCCTCGACCAGCAGGTCAAGCTGCGTGGCTTCCGCATCGAGCCGGAGGAAAT 255

Qy        2817 CGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAA 2876

Db         254 CCAGGCACGCCTGCTGGCGCAGCCGGGGGTGGCCCAGGCGGTAGTGGTGATCCGCGAAGG 195

Qy        2877 CGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCAC 2924

Db         194 CGTGGCCGGCAGCCAGTTGGTCGGCTACTACACCGGCGCTGTCGGCGC 147

Search completed: October  3, 2003, 17:48:03
Job time : 8366.86 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 03:31:13 ; Search time 973.777 Seconds
(without alignments)
11088.586 Million cell updates/sec

Title: US-10-014-717-1_COPY_11872_16104
Perfect score: 4233
Sequence: 1 atgacgatcaatcagcttct..........agggcaggagacgtagctaa 4233

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4233	100.0	68750	13	US-10-014-717-1	Sequence 1, Appli
	2	543.4	12.8	4305	12	US-10-238-075-940	Sequence 940, App
c	3	543.4	12.8	48715	12	US-10-238-075-921	Sequence 921, App
c	4	543.4	12.8	50538	12	US-10-085-959-57	Sequence 57, Appl
	5	390.4	9.2	32159	12	US-10-085-959-54	Sequence 54, Appl
	6	360.6	8.5	135638	12	US-10-314-657-1	Sequence 1, Appli
	7	305.6	7.2	1416	9	US-09-815-242-7601	Sequence 7601, Ap
	8	296.8	7.0	1875	14	US-10-156-761-3143	Sequence 3143, Ap
c	9	296.8	7.0	9025608	14	US-10-156-761-1	Sequence 1, Appli
	10	270.4	6.4	5304	14	US-10-156-761-3184	Sequence 3184, Ap
	11	261.4	6.2	13029	9	US-09-815-242-4052	Sequence 4052, Ap
	12	246.2	5.8	11058	14	US-10-156-761-3629	Sequence 3629, Ap
	13	241.6	5.7	77536	12	US-09-940-316B-1	Sequence 1, Appli
	14	239	5.6	88421	10	US-09-976-059-1	Sequence 1, Appli
	15	233.4	5.5	5127	14	US-10-156-761-3630	Sequence 3630, Ap
c	16	232.6	5.5	5307	12	US-10-238-075-1078	Sequence 1078, Ap
	17	230.2	5.4	1623	14	US-10-156-761-848	Sequence 848, App
	18	230.2	5.4	9025608	14	US-10-156-761-1	Sequence 1, Appli
	19	221.8	5.2	7347	9	US-09-815-242-7773	Sequence 7773, Ap
	20	218	5.2	2934	14	US-10-156-761-3186	Sequence 3186, Ap
c	21	203.8	4.8	135638	12	US-10-314-657-1	Sequence 1, Appli
	22	202.2	4.8	1706	12	US-10-238-075-1079	Sequence 1079, Ap
	23	194.6	4.6	3048	14	US-10-156-761-3146	Sequence 3146, Ap
	24	190.2	4.5	3789	14	US-10-156-761-3185	Sequence 3185, Ap
	25	188	4.4	3348	12	US-10-238-075-1081	Sequence 1081, Ap
	26	177.4	4.2	4341	14	US-10-166-087-45	Sequence 45, Appl
	27	177.4	4.2	32539	14	US-10-166-087-1	Sequence 1, Appli
	28	161.8	3.8	1923	14	US-10-156-761-836	Sequence 836, App
c	29	158.2	3.7	1178	10	US-09-924-256A-91	Sequence 91, Appl
	30	157.4	3.7	1545	14	US-10-156-761-1239	Sequence 1239, Ap
	31	155.2	3.7	3471	10	US-09-974-300-2167	Sequence 2167, Ap
	32	155	3.7	1200	10	US-09-924-256A-85	Sequence 85, Appl
c	33	150.8	3.6	1204	10	US-09-924-256A-87	Sequence 87, Appl
	34	148.4	3.5	1803	14	US-10-166-087-43	Sequence 43, Appl
c	35	137.4	3.2	1172	10	US-09-924-256A-17	Sequence 17, Appl
	36	136.6	3.2	3798	10	US-09-974-300-2175	Sequence 2175, Ap
	37	136.6	3.2	3798	10	US-09-974-300-2193	Sequence 2193, Ap
	38	132.8	3.1	1590	14	US-10-156-761-841	Sequence 841, App
	39	132	3.1	3885	9	US-09-815-242-9626	Sequence 9626, Ap
	40	126.8	3.0	3234	10	US-09-974-300-2168	Sequence 2168, Ap
	41	126	3.0	1722	14	US-10-156-761-834	Sequence 834, App
	42	122.8	2.9	4320	14	US-10-156-761-595	Sequence 595, App
	43	120.8	2.9	7158	10	US-09-974-300-2171	Sequence 2171, Ap
	44	119.4	2.8	1806	14	US-10-156-761-827	Sequence 827, App
	45	118.8	2.8	3882	9	US-09-815-242-5982	Sequence 5982, Ap

ALIGNMENTS

RESULT 1
US-10-014-717-1
; Sequence 1, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-014-717-1

```
  Query Match             100.0%;  Score 4233;  DB 13;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy      181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy      241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy      301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy      361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy      421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy      481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy      541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy      601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy      661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy      721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy      781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy      841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy      901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy      961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy     1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy     1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy     1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy     1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy     1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy     1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy     1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy     1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy     1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy     1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy     1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy     1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy     1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy     1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy     1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy     1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy     1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy     2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy     2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy     2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy     2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy     2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211
```

```
Qy       2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy       2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy       2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy       2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy       2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy       2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy       2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy       2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy       2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy       2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy       2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy       3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy       3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy       3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy       3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy       3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy       3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy       3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||||||||||||||||||||||||||||||||
Db      16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

```
RESULT 2
US-10-238-075-940
; Sequence 940, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
;  APPLICANT: I.N.S.E.R.M.
;  TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
;  TITLE OF INVENTION:  E.coli, and biological uses of these polynucleotides and of thei
;  FILE REFERENCE: BLANDINE
;  CURRENT APPLICATION NUMBER: US/10/238,075
;  CURRENT FILING DATE:  2002-09-10
;  PRIOR APPLICATION NUMBER: 0003145
;  PRIOR FILING DATE: 2000-03-10
;  NUMBER OF SEQ ID NOS: 1576
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 940
;   LENGTH: 4305
```

```
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-10-238-075-940

  Query Match             12.8%;  Score 543.4;  DB 12;  Length 4305;
  Best Local Similarity   48.5%;  Pred. No. 6.8e-132;
  Matches 1933;  Conservative    0;  Mismatches 1961;  Indels   93;  Gaps   12;

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
Db        241 ATAGTGCACGACCCTGAGCATCGTCATCAACCGTTCCCATTGACGGATGTTCAGCGCGCC 300

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
Db        301 TACTGGCTTGGACG---ACAGACGGGTGCTACCTCGATCGCGACCCACATCTACCATGAA 357

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
Db        358 TTTGACGTAGAACACTTTAATGTTACGCGTTTTACCCATGCGGTGAATGCGCTGATCGCT 417

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
Db        418 CGCCATGAAATGCTACGTGCGCGGGTACTCCCCGACGGTACTCAGCAGATTCTGGCGCAA 477

Qy        421 GTCGACGC---CGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAA 477
Db        478 GTGCCGGCGTATCAGTTAGAGCAGCGCGATCTGAGTGCTTTGTCCCCTAACGCACGAAAC 537

Qy        478 GCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCT 537
Db        538 GATGCCTTGATGGCGATCCGCGATCGGCTGTCGCATCATGTGCATCCCGCAGATCGTTGG 597

Qy        538 CCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGT 597
Db        598 CCGCTGTTTGATTTCAGTTATTCGGCTTGCACGGCGCAACATGGCCGCTTGCATTTCAGT 657

Qy        598 ATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGC 657
Db        658 CTCGATCTGCTGATTGCCGATGCTCTGAGTATGCGCACGCTACAGCAGGAGTTGATGATG 717

Qy        658 TTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTA 717
Db        718 CTGTACCGTGAGCCCCATGTGTCACTGCCGTTGCTACCGTTCTCTTTTCGTGACTACGTG 777

Qy        718 CTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAG 777
Db        778 CAGGCGCTGTTGGTAGAGCAGGCGAGTGAAGCCTATGCACGCGATCAGGCCTATTGGCAA 837

Qy        778 CGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACC 837
Db        838 CGGGCGCTGCCGCAGCTGTATGGCCCACCAACGCTGCCCGTACAGGGCGATTTGGCGCAA 897

Qy        838 CTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGA 897
Db        898 CTGTCTGCGATCAGGTTCGTACGTCGCCGTCATCGGCTGTCAGCCCACAACTGGGGAGTG 957

Qy        898 TTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTT 957
Db        958 CTGAGCGCGCTGGCCCAACGCACACGTATCACCAAGACGGCATTGTTGTTGACAGTCTTT 1017

Qy        958 TCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTC 1017
Db       1018 AGCCAAGTGCTGGCACGTTGGAGCCTTAGCCCGACGTTTACGCTCAATCTGACGTTGTTC 1077

Qy       1018 AACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTC 1077
Db       1078 AACCGCCCGCAGGGTTACCCCAACGCAGAGGCAGTTATTGGTGATTTTACCGCTGTCAGC 1137

Qy       1078 CTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAA 1137
Db       1138 TTGCTGAATGTTTGTTACGACAGCCAGCACTCTTATGCCCACAACGCTCAGCGTATTCAG 1197

Qy       1138 GAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTAT---------CGAGGTC 1188
Db       1198 GTGCAACTGTGGGAAGATCTCGAACATCGTCGTTTCAGTGGGATCCGCGCCAGCGAGGCG 1257

Qy       1189 CAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTC 1248
Db       1258 CTGATCCATAGCGGTCGTTTCCATGCGCCGATGCCGGTGGTATTCACTAGTATGTTGGAT 1317

Qy       1249 ACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCG 1308
Db       1318 ATCGACGGGGAGACGACTGCGCAAGACCCTCGGGACACAACCCGTTTTACTCTGTGTCCG 1377

Qy       1309 GTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGAT 1368
Db       1378 GACGCCAATATTACCCAAACACCGCAGGTGTGGCTCGATCACCAGGTGATCGAGTTGGCT 1437

Qy       1369 GGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGAC 1428
Db       1438 GGGGAGTTGCATTTCAACTGGGACGCGGTCGAGCAACTGTTTGATACCACGCTGCTGGAT 1497

Qy       1429 GACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAA 1488
Db       1498 CAGATGTTTGGTGCTTATTGTCATGCGCTGCAGGCGCTGGTTGCCATGCC---------G 1548

Qy       1489 CAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACC 1548
Db       1549 CAAAGTTGGTGGGGGGTAAATAGTTCTCTGGCGCTGCCCACCGTTAGTGCACCGGTCACG 1608

Qy       1549 AACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTG 1608
Db       1609 CAGGCTCCTGCACCTACGGCCTTGTTGCACCATGGATTACTGCGTCAGGCAGCACTGACG 1668

Qy       1609 CCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGC 1668
Db       1669 CCACAGGAAACTGCGCTGATCAGTCCTATCCGTGAATTGACCTATCGCCAACTGTCGACG 1728

Qy       1669 CGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTC 1728
Db       1729 GCGGCGGATCATGTGGCCCGCGCCCTGTTAGCGCTGGGCGTGCAGCATGGCGACCGCGTG 1788

Qy       1729 GCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCA 1788
Db       1789 GCGGTGGTGATGGAAAAAGGCTGGCAGCAGATTGCCGCCGTACACGGCATTTTACGACTG 1848

Qy       1789 GGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTC 1848
Db       1849 GGTGCGGTCTATCTGCCAGTGGATCCGGTGCTACCGCCACAGCGTCGCCAGCTTTTGCTG 1908

Qy       1849 GATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGG 1908
Db       1909 ACGGTGGGCGAGGTGCGGGTACAAGTAACGCAGCCGGGTCTCACGCAATTGGAGCCGTCG 1968

Qy       1909 CCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAG 1968
Db       1969 CTGCCCG------------TGCTGATCATCGACGACGGAATGCTGGACACGCCTGCTGCG 2016

Qy       1969 CCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGA 2028
Db       2017 CCGTTGCCTGAAGTGGCTGGGGATGTCACGGATCTGGCCTATATCATTTTCACTTCCGGC 2076

Qy       2029 TCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTG 2088
Db       2077 TCCACCGGTACCCCGAAAGGAGTGATGATCGACCACCGTGCGGCCATGAACACGCTGGAA 2136

Qy       2089 GACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTG 2148
Db       2137 GACATCAACGAACGCTTTGGCCTCAATGCGCAGGATAGGGTGTTCGGGCTGTCATCATTG 2196

Qy       2149 AGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTG 2208
Db       2197 AGCTTTGACCTGTCGGTTTACGATGCCTTTGCGCCTTTTATGGTGGGTGCAGCGCTGGTA 2256

Qy       2209 GTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAG 2268
```

```
Db   2257 CTGCCGGAAGCAGGACGGGAAAAAGATCCGCGTCATTGGCAGACAGTTATGGCACACGGT 2316

Qy   2269 AAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTGAG 2328
Db   2317 CATGTAAGCGTCTGGAATGCAGTGCCCGCACTGATGCAGATGCTGTGCG---AATACCAC 2373

Qy   2329 GGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATC 2388
Db   2374 AGCGGCGATCGGATGAGTTATCCGACGTTGCGTCTGGCACTGTTGAGCGGCGACTGGATC 2433

Qy   2389 CCGGTGGCCTGCCTGGCGA---GCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGC 2445
Db   2434 CCGCTAACGTTACCGGAGCAGATGCGCGAGCGGCTCAATGAAACGATGGACATCATCAGT 2493

Qy   2446 CTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGAC 2505
Db   2494 CTGGGTGGAGCGACCGAGTGCGCCATCTGGTCGGTCTACTACCCGATAGGTGAGGTGGAA 2553

Qy   2506 CTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTC 2565
Db   2554 TCGACGTGGACCAGTATTCCCTACGGTCGGGGCCTGCGCAACCAGCCAGTATACGTGCTA 2613

Qy   2566 GATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTC 2625
Db   2614 AATGCGCAACTGGAGGAATGTCCGGTCGGGGTGGAAGGAGAGATTTGCATTGGCGGGATG 2673

Qy   2626 GGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCAC 2685
Db   2674 GGGCTGGCACAAGGCTACCTGAACGACGCAGAGAAAACGGCGGCGAGCTTTGTCTGGCGC 2733

Qy   2686 CCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGA 2745
Db   2734 GAAGCGAGTGGTGAGCGAATTTACCGCACTGGGGATCGCGGGCGCTACTTTGCTGACGGG 2793

Qy   2746 AACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAG 2805
Db   2794 CAAGTCGCCTTTTTGGGGCGCAACGATACCCAAGTGAAGGTGAATGGTTACCGTATCGAA 2853

Qy   2806 CTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTG 2865
Db   2854 CTGGGGGAAATCGAGCGCTGCATTGCGCGACATCCCGATGTGGAGCAGTCAGTGGTGGTG 2913

Qy   2866 CCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACA 2925
Db   2914 GCAGTGGGTAA---------------------------------TTCTCAACATCGTCGG 2940

Qy   2926 CGGAGACGCGCTGCCGAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCA 2985
Db   2941 CTGGTCGCTTTTGCCAAACTGCACGATCGCCACCAGGCGCAGGCATTGCAAGCTAAGGAA 3000

Qy   2986 CACGCCGCCGAAGCGGACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGA 3045
Db   3001 ------GCGGAGGCGGCGGCACTGGCGCAGGGTATTATTGTGAATCCGGCACAGCGTCTA 3054

Qy   3046 CACGGACTCCGGAGGGACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCG 3105
Db   3055 GCGTTCAAACTCAAGGAGCCACATATTCGCGCGCTGGATGGTCTGGGCATTGCACTGACG 3114

Qy   3106 CGGGAGGCGGGGCTGGACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCC 3165
Db   3115 GCACCGGCGGATAGCACACGTTACATCAAACGCCGCAGCTATCGTCATTTCAGCGCGCAA 3174

Qy   3166 CCGATTCCGTTTGTTGAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGAC 3225
Db   3175 AAAACCACGCTGGCACAGTTGGGGCAATTGCTGTCGGGCTTGGGGCAGATGCGTCTACCC 3234

Qy   3226 GGCGCGACCCTTCCCAAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACC 3285
Db   3235 GGTCTACCTTTTGCCAAGTATGCCTATGCGTCCGCCGGGGGCTATACCCGGTGCAAACC 3294

Qy   3286 TACGCGTATGTCAAATCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCAC 3345
Db   3295 TACGTGTACCTGCATCCAGACAAGATCGAAGAGGGAGTATCCGGTATTTACTACTTCGAC 3354

Qy   3346 CCGTTCGAGCACCGTTTGCTGAAGCT---CTCCGATCACGGGATCGAGCGCGGAGCGCAC 3402
Db   3355 CCGCGACAGAGCTGTCTTATGCCGGTAGCACCAGAAGTCGAGCTGAACAGTGGTTTTCAT 3414

Qy   3403 GTTCGGCAAAACTTCGACGTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGG 3462
Db   3415 GCCGGACCTAATCAGTCTATTGCCGATCGGGCGGCATTCACGCTGTTTATGGTGGCTGAT 3474

Qy   3463 ATCGACGCCATCGAGTCGCTGTATGG---ATCGTCGTCGCGAGAATTTTGCCTGCTGGAG 3519
Db   3475 ATGGCGGTGATCTCGCCATTCTATGGGCAAGAGGCAGCTTGGCACTTCTCGGTGATGGAA 3534

Qy   3520 GCCGGATATATGGCGCAGCTCCTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGT 3579
Db   3535 GCAGGTACTCTCTGCCATTTACTGGAAGAAGATGCGCCGCGCTACGGATTGGGGCTGTGC 3594

Qy   3580 CCGGTGGGGCAATTCAATTTTGAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGAC 3639
Db   3595 CAACTTGGGATGGCAGACTTTTCCGCTGTGGCATCGCATTTTCAATTGTCGCCACATCAT 3654

Qy   3640 GTTTACGTGCACGGCATGCTGGGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACG 3699
Db   3655 CGCTATGTCCATTGCACCGTGGGGGGCGCGATAGGGCAAGAGGCGGCAAGTGCTGCAGCA 3714

Qy   3700 CTCGGTCAGGATTCCTCACCGAGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAG 3759
Db   3715 TTGCTGCGCGATTTCTCCACCTATGAGAAACCGAAGGAAACCGCTGCGCCGCTGGACATG 3774

Qy   3760 CAGCACTTCGCCGATATGCTTCGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTG 3819
Db   3775 CAGAGCTACAAAGATGCCATGCTGCGCGGCCTGCGTCAGCAACTGCCTGACTATATGGTG 3834

Qy   3820 CCTACAGTCTTCGTGGAGCTCGATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGT 3879
Db   3835 CCGAGTGATCTGATGTTAGCGACCGATTTTCCGTTAACCGCTAACGGCAAGCTAGATCGG 3894

Qy   3880 AAGGCCCTGCGCGAGCGGAAGGATACCTCGTCGCCGCGGCAT------TCGGGGCACACG 3933
Db   3895 CAAAAATTACAGCTGCAGGGCGAACAAATTGCCCACCAGCGTGACGGCGTGGGTCCAATC 3954

Qy   3934 GCGCCACGGGACGCCTTGGAGGAGATCCTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTG 3993
Db   3955 CAGGTGGACAGTGCGTTACAACAGCGGCTGGTGGCGCTCTGGCAGGAGGTACTCGGCGTG 4014

Qy   3994 GAGGTGGTCGGGCTCCAGCAGAGCTTCGTCGATCTTGGTGCGACATCGATTCACATCGTT 4053
Db   4015 AGCCACGTGTCGGCCGAAGACGATTTCTTCTCGCTGGGGGGCAGTTCTATAGAATTGGTG 4074

Qy   4054 CGCATGAGGAGCCTGTTGCAGAAGAGGCTGGATAGGGAGATCGCCATCACCGAGTTGTTC 4113
Db   4075 CGTATTCAGCAGGCACTTGAGGCGATTATCGGGCAGGAGATTCCCATTGTCGATCTGTTC 4134

Qy   4114 CAGTACCCGAACCTCGGCTCGCTGGCG 4140
Db   4135 CGTCTGCCAACCATCGCGGATGTGGCG 4161
```

RESULT 3
US-10-238-075-921/c
; Sequence 921, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of thei
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1

```
; SEQ ID NO 921
;   LENGTH: 48715
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-10-238-075-921

  Query Match             12.8%;  Score 543.4;  DB 12;  Length 48715;
  Best Local Similarity   48.5%;  Pred. No. 1.1e-131;
  Matches 1933;  Conservative    0;  Mismatches 1961;  Indels   93;  Gaps   12;

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
Db      18373 ATAGTGCACGACCCTGAGCATCGTCATCAACCGTTCCCATTGACGGATGTTCAGCGCGCC 18314

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
Db      18313 TACTGGCTTGGACG---ACAGACGGGTGCTACCTCGATCGCGACCCACATCTACCATGAA 18257

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
Db      18256 TTTGACGTAGAACACTTTAATGTTACGCGTTTTACCCATGCGGTGAATGCGCTGATCGCT 18197

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
Db      18196 CGCCATGAAATGCTACGTGCGCGGGTACTCCCCGACGGTACTCAGCAGATTCTGGCGCAA 18137

Qy        421 GTCGACGC---CGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAA 477
Db      18136 GTGCCGGCGTATCAGTTAGAGCAGCGCGATCTGAGTGCTTTGTCCCCTAACGCACGAAAC 18077

Qy        478 GCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCT 537
Db      18076 GATGCCTTGATGGCGATCCGCGATCGGCTGTCGCATCATGTGCATCCCGCAGATCGTTGG 18017

Qy        538 CCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGT 597
Db      18016 CCGCTGTTTGATTTCAGTTATTCGGCTTGCACGGCGCAACATGGCCGCTTGCATTTCAGT 17957

Qy        598 ATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGC 657
Db      17956 CTCGATCTGCTGATTGCCGATGCTCTGAGTATGCGCACGCTACAGCAGGAGTTGATGATG 17897

Qy        658 TTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTA 717
Db      17896 CTGTACCGTGAGCCCCATGTGTCACTGCCGTTGCTACCGTTCTCTTTTCGTGACTACGTG 17837

Qy        718 CTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAG 777
Db      17836 CAGGCGCTGTTGGTAGAGCAGGCGAGTGAAGCCTATGCACGCGATCAGGCCTATTGGCAA 17777

Qy        778 CGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACC 837
Db      17776 CGGGCGCTGCCGCAGCTGTATGGCCCACCAACGCTGCCCGTACAGGGCGATTTGGCGCAA 17717

Qy        838 CTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGA 897
Db      17716 CTGTCTGCGATCAGGTTCGTACGTCGCCGTCATCGGCTGTCAGCCCACAACTGGGGAGTG 17657

Qy        898 TTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTT 957
Db      17656 CTGAGCGCGCTGGCCCAACGCACACGTATCACCAAGACGGCATTGTTGTTGACAGTCTTT 17597

Qy        958 TCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTC 1017
Db      17596 AGCCAAGTGCTGGCACGTTGGAGCCTTAGCCCGACGTTTACGCTCAATCTGACGTTGTTC 17537

Qy       1018 AACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTC 1077
Db      17536 AACCGCCCGCAGGGTTACCCCAACGCAGAGGCAGTTATTGGTGATTTTACCGCTGTCAGC 17477

Qy       1078 CTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAA 1137
Db      17476 TTGCTGAATGTTTGTTACGACAGCCAGCACTCTTATGCCCACAACGCTCAGCGTATTCAG 17417

Qy       1138 GAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTAT---------CGAGGTC 1188
Db      17416 GTGCAACTGTGGGAAGATCTCGAACATCGTCGTTTCAGTGGGATCCGCGCCAGCGAGGCG 17357

Qy       1189 CAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTC 1248
Db      17356 CTGATCCATAGCGGTCGTTTCCATGCGCCGATGCCGGTGGTATTCACTAGTATGTTGGAT 17297

Qy       1249 ACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCG 1308
Db      17296 ATCGACGGGGAGACGACTGCGCAAGACCCTCGGGACACAACCCGTTTACTCTGTGTCCG 17237

Qy       1309 GTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGAT 1368
Db      17236 GACGCCAATATTACCCAAACACCGCAGGTGTGGCTCGATCACCAGGTGATCGAGTTGGCT 17177

Qy       1369 GGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGAC 1428
Db      17176 GGGGAGTTGCATTTCAACTGGGACGCGGTCGAGCAACTGTTTGATACCACGCTGCTGGAT 17117

Qy       1429 GACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAA 1488
Db      17116 CAGATGTTTGGTGCTTATTGTCATGCGCTGCAGGCGCTGGTTGCCATGCC---------G 17066

Qy       1489 CAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACC 1548
Db      17065 CAAAGTTGGTGGGGGGTAAATAGTTCTCTGGCGCTGCCCACCGTTAGTGCACCGGTCACG 17006

Qy       1549 AACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTG 1608
Db      17005 CAGGCTCCTGCACCTACGGCCTTGTTGCACCATGGATTACTGCGTCAGGCAGCACTGACG 16946

Qy       1609 CCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGC 1668
Db      16945 CCACAGGAAACTGCGCTGATCAGTCCTATCCGTGAATTGACCTATCGCCAACTGTCGACG 16886

Qy       1669 CGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTC 1728
Db      16885 GCGGCGGATCATGTGGCCCGCGCCCTGTTAGCGCTGGGCGTGCAGCATGGCGACCGCGTG 16826

Qy       1729 GCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCA 1788
Db      16825 GCGGTGGTGATGGAAAAAGGCTGGCAGCAGATTGCCGCCGTACACGGCATTTTACGACTG 16766

Qy       1789 GGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTC 1848
Db      16765 GGTGCGGTCTATCTGCCAGTGGATCCGGTGCTACCGCCACAGCGTCGCCAGCTTTTGCTG 16706

Qy       1849 GATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGG 1908
Db      16705 ACGGTGGGCGAGGTGCGGGTACAAGTAACGCAGCCGGGTCTCACGCAATTGGAGCCGTCG 16646

Qy       1909 CCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAG 1968
Db      16645 CTGCCCG-----------TGCTGATCATCGACGACGGAATGCTGGACACGCCTGCTGCG 16598

Qy       1969 CCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGA 2028
Db      16597 CCGTTGCCTGAAGTGGCTGGGGATGTCACGGATCTGGCCTATATCATTTTCACTTCCGGC 16538

Qy       2029 TCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTG 2088
Db      16537 TCCACCGGTACCCCGAAAGGAGTGATGATCGACCACCGTGCGGCCATGAACACGCTGGAA 16478

Qy       2089 GACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTG 2148
Db      16477 GACATCAACGAACGCTTTGGCCTCAATGCGCAGGATAGGGTGTTCGGGCTGTCATCATTG 16418

Qy       2149 AGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTG 2208
Db      16417 AGCTTTGACCTGTCGGTTTACGATGCCTTTGCGCCTTTTATGGTGGGTGCAGCGCTGGTA 16358
```

```
Qy       2209 GTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGCAGAGTTGATCGAACGAGAG 2268
Db      16357 CTGCCGGAAGCAGGACGGGAAAAAGATCCGCGTCATTGGCAGACAGTTATGGCACACGGT 16298

Qy       2269 AAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTGAG 2328
Db      16297 CATGTAAGCGTCTGGAATGCAGTGCCCGCACTGATGCAGATGCTGTGCG---AATACCAC 16241

Qy       2329 GGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATC 2388
Db      16240 AGCGGCGATCGGATGAGTTATCCGACGTTGCGTCTGGCACTGTTGAGCGGCGACTGGATC 16181

Qy       2389 CCGGTGGGCCTGCCTGGCGA---GCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGC 2445
Db      16180 CCGCTAACGTTACCGGAGCAGATGCGCGAGCGGCTCAATGAAACGATGGACATCATCAGT 16121

Qy       2446 CTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGAC 2505
Db      16120 CTGGGTGGAGCGACCGAGTGCGCCATCTGGTCGGTCTACTACCCGATAGGTGAGGTGGAA 16061

Qy       2506 CTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTC 2565
Db      16060 TCGACGTGGACCAGTATTCCCTACGGTCGGGGCCTGCGCAACCAGCCAGTATACGTGCTA 16001

Qy       2566 GATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTC 2625
Db      16000 AATGCGCAACTGGAGGAATGTCCGGTCGGGGTGGAAGGAGAGATTTGCATTGGCGGGATG 15941

Qy       2626 GGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCAC 2685
Db      15940 GGGCTGGCACAAGGCTACCTGAACGACGCAGAGAAAACGGCGGCGAGCTTTGTCTGGCGC 15881

Qy       2686 CCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGA 2745
Db      15880 GAAGCGAGTGGTGAGCGAATTTACCGCACTGGGGATCGCGGGCGCTACTTTGCTGACGGG 15821

Qy       2746 AACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAG 2805
Db      15820 CAAGTCGCCTTTTTGGGGCGCAACGATACCCAAGTGAAGGTGAATGGTTACCGTATCGAA 15761

Qy       2806 CTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTG 2865
Db      15760 CTGGGGGAAATCGAGCGCTGCATTGCGCGACATCCCGATGTGGAGCAGTCAGTGGTGGTG 15701

Qy       2866 CCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACA 2925
Db      15700 GCAGTGGGTAA--------------------------------TTCTCAACATCGTCGG 15674

Qy       2926 CGGAGACGCGCTGCCGAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCA 2985
Db      15673 CTGGTCGCTTTTGCCAAACTGCACGATCGCCACCAGGCGCAGGCATTGCAAGCTAAGGAA 15614

Qy       2986 CACGCCGCCGAAGCGGACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGA 3045
Db      15613 ------GCGGAGGCGGCGGCACTGGCGCAGGGTATTATTGTGAATCCGGCACAGCGTCTA 15560

Qy       3046 CACGGACTCCGGAGGGACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCG 3105
Db      15559 GCGTTCAAACTCAAGGAGCCACATATTCGCGCGCTGGATGGTCTGGGCATTGCACTGACG 15500

Qy       3106 CGGGAGGCGGGGCTGGACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCC 3165
Db      15499 GCACCGGCGGATAGCACACGTTACATCAAACGCCGCAGCTATCGTCATTTCAGCGCGCAA 15440

Qy       3166 CCGATTCCGTTTGTTGAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGAC 3225
Db      15439 AAAACCACGCTGGCACAGTTGGGGCAATTGCTGTCGGGCTTGGGGCAGATGCGTCTACCC 15380

Qy       3226 GGCGCGACCCTTCCCAAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACC 3285
Db      15379 GGTCTACCTTTTGCCAAGTATGCCTATGCGTCCGCCGGGGGGCTATACCCGGTGCAAACC 15320

Qy       3286 TACGCGTATGTCAAATCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCAC 3345
Db      15319 TACGTGTACCTGCATCCAGACAAGATCGAAGAGGGAGTATCCGGTATTTACTACTTCGAC 15260

Qy       3346 CCGTTCGAGCACCGTTTGCTGAAGCT---CTCCGATCACGGGATCGAGCGCGGAGCGCAC 3402
Db      15259 CCGCGACAGAGCTGTCTTATGCCGGTAGCACCAGAAGTCGAGCTGAACAGTGGTTTTCAT 15200

Qy       3403 GTTCGGCAAAACTTCGACGTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGG 3462
Db      15199 GCCGGACCTAATCAGTCTATTGCCGATCGGGCGGCATTCACGCTGTTTATGGTGGCTGAT 15140

Qy       3463 ATCGACGCCATCGAGTCGCTGTATGG---ATCGTCGTCGCGAGAATTTGCCTGCTGGAG 3519
Db      15139 ATGGCGGTGATCTCGCCATTCTATGGGCAAGAGGCAGCTTGGCACTTCTCGGTGATGGAA 15080

Qy       3520 GCCGGATATATGGCGCAGCTCCTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGT 3579
Db      15079 GCAGGTACTCTCTGCCATTTACTGGAAGAAGATGCGCCGCGCTACGGATTGGGGCTGTGC 15020

Qy       3580 CCGGTGGGGCAATTCAATTTTGAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGAC 3639
Db      15019 CAACTTGGGATGGCAGACTTTTCCGCTGTGGCATCGCATTTTCAATTGTCGCCACATCAT 14960

Qy       3640 GTTTACGTGCACGGCATGCTGGGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACG 3699
Db      14959 CGCTATGTCCATTGCACCGTGGGGGGCGCGATAGGGCAAGAGGCGGCAAGTGCTGCAGCA 14900

Qy       3700 CTCGGTCAGGATTCCTCACCGAGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAG 3759
Db      14899 TTGCTGCGCGATTTCTCCACCTATGAGAAACCGAAGGAAACCGCTGCGCCGCTGGACATG 14840

Qy       3760 CAGCACTTCGCCGATATGCTTCGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTG 3819
Db      14839 CAGAGCTACAAAGATGCCATGCTGCGCGGCCTGCGTCAGCAACTGCCTGACTATATGGTG 14780

Qy       3820 CCTACAGTCTTCGTGGAGCTCGATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGT 3879
Db      14779 CCGAGTGATCTGATGTTAGCGACCGATTTTCCGTTAACCGCTAACGGCAAGCTAGATCGG 14720

Qy       3880 AAGGCCCTGCGCGAGCGGAAGGATACCTCGTCGCCGCGGCAT------TCGGGGCACACG 3933
Db      14719 CAAAAATTACAGCTGCAGGGCGAACAAATTGCCCACCAGCGTGACGGCGTGGTCCAATC 14660

Qy       3934 GCGCCACGGGACGCCTTGGAGGAGATCCTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTG 3993
Db      14659 CAGGTGGACAGTGCGTTACAACAGCGGCTGGTGGCGCTCTGGCAGGAGGTACTCGGCGTG 14600

Qy       3994 GAGGTGGTCGGGCTCCAGCAGAGCTTCGTCGATCTTGGTGCGACATCGATTCACATCGTT 4053
Db      14599 AGCCACGTGTCGGCCGAAGACGATTTCTTCTCGCTGGGGGGCAGTTCTATAGAATGGTG 14540

Qy       4054 CGCATGAGGAGCCTGTTGCAGAAGAGGCTGGATAGGGAGATCGCCATCACCGAGTTGTTC 4113
Db      14539 CGTATTCAGCAGGCACTTGAGGCGATTATCGGGCAGGAGATTCCCATTGTCGATCTGTTC 14480

Qy       4114 CAGTACCCGAACCTCGGCTCGCTGGCG 4140
Db      14479 CGTCTGCCAACCATCGCGGATGTGGCG 14453
```

RESULT 4
US-10-085-959-57/c
; Sequence 57, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1el Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412

```
;  PRIOR FILING DATE: 2000-10-19
;  NUMBER OF SEQ ID NOS: 255
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
;   LENGTH: 50538
;   TYPE: DNA
;   ORGANISM: Escherichia coli
US-10-085-959-57

  Query Match             12.8%;  Score 543.4;  DB 12;  Length 50538;
  Best Local Similarity   48.5%;  Pred. No. 1.1e-131;
  Matches 1933;  Conservative    0;  Mismatches 1961;  Indels   93;  Gaps   12;

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
Db      20201 ATAGTGCACGACCCTGAGCATCGTCATCAACCGTTCCCATTGACGGATGTTCAGCGCGCC 20142

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
Db      20141 TACTGGCTTGGACG---ACAGACGGGTGCTACCTCGATCGCGACCCACATCTACCATGAA 20085

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
Db      20084 TTTGACGTAGAACACTTTAATGTTACGCGTTTTACCCATGCGGTGAATGCGCTGATCGCT 20025

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
Db      20024 CGCCATGAAATGCTACGTGCGCGGGTACTCCCCGACGGTACTCAGCAGATTCTGGCGCAA 19965

Qy        421 GTCGACGC---CGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAA 477
Db      19964 GTGCCGGCGTATCAGTTAGAGCAGCGCGATCTGAGTGCTTTGTCCCCTAACGCACGAAAC 19905

Qy        478 GCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCT 537
Db      19904 GATGCCTTGATGGCGATCCGCGATCGGCTGTCGCATCATGTGCATCCCGCAGATCGTTGG 19845

Qy.       538 CCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGT 597
Db      19844 CCGCTGTTTGATTTCAGTTATTCGGCTTGCACGGCGCAACATGGCCGCTTGCATTTCAGT 19785

Qy        598 ATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGC 657
Db      19784 CTCGATCTGCTGATTGCCGATGCTCTGAGTATGCGCACGCTACAGCAGGAGTTGATGATG 19725

Qy        658 TTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTA 717
Db      19724 CTGTACCGTGAGCCCCATGTGTCACTGCCGTTGCTACCGTTCTCTTTTCGTGACTACGTG 19665

Qy        718 CTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAG 777
Db      19664 CAGGCGCTGTTGGTAGAGCAGGCGAGTGAAGCCTATGCACGCGATCAGGCCTATTGGCAA 19605

Qy        778 CGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACC 837
Db      19604 CGGGCGCTGCCGCAGCTGTATGGCCCACCAACGCTGCCCGTACAGGGCGATTTGGCGCAA 19545

Qy        838 CTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGA 897
Db      19544 CTGTCTGCGATCAGGTTCGTACGTCGCCGTCATCGGCTGTCAGCCCACAACTGGGGAGTG 19485

Qy        898 TTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTT 957
Db      19484 CTGAGCGCGCTGGCCCAACGCACACGTATCACCAAGACGGCATTGTTGTTGACAGTCTTT 19425

Qy        958 TCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTC 1017
Db      19424 AGCCAAGTGCTGGCACGTTGGAGCCTTAGCCCGACGTTTACGCTCAATCTGACGTTGTTC 19365

Qy       1018 AACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTC 1077
Db      19364 AACCGCCCGCAGGGTTACCCCAACGCAGAGGCAGTTATTGGTGATTTTACCGCTGTCAGC 19305

Qy       1078 CTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAA 1137
Db      19304 TTGCTGAATGTTTGTTACGACAGCCAGCACTCTTATGCCCACAACGCTCAGCGTATTCAG 19245

Qy       1138 GAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTAT---------CGAGGTC 1188
Db      19244 GTGCAACTGTGGGAAGATCTCGAACATCGTCGTTTCAGTGGGATCCGCGCCAGCGAGGCG 19185

Qy       1189 CAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTC 1248
Db      19184 CTGATCCATAGCGGTCGTTTCCATGCGCCGATGCCGGTGGTATTCACTAGTATGTTGGAT 19125

Qy       1249 ACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCG 1308
Db      19124 ATCGACGGGGAGACGACTGCGCAAGACCCTCGGGACACAACCCGTTTTACTCTGTGTCCG 19065

Qy       1309 GTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGAT 1368
Db      19064 GACGCCAATATTACCCAAACACCGCAGGTGTGGCTCGATCACCAGGTGATCGAGTTGGCT 19005

Qy       1369 GGGGACCTCGTCCTCGCGTGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGAC 1428
Db      19004 GGGGAGTTGCATTTCAACTGGGACGCGGTCGAGCAACTGTTTGATACCACGCTGCTGGAT 18945

Qy       1429 GACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAA 1488
Db      18944 CAGATGTTTGGTGCTTATTGTCATGCGCTGCAGGCGCTGGTTGCCATGCC---------G 18894

Qy       1489 CAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACC 1548
Db      18893 CAAAGTTGGTGGGGGGTAAATAGTTCTCTGGCGCTGCCCACCGTTAGTGCACCGGTCACG 18834

Qy       1549 AACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTG 1608
Db      18833 CAGGCTCCTGCACCTACGGCCTTGTTGCACCATGGATTACTGCGTCAGGCAGCACTGACG 18774

Qy       1609 CCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGC 1668
Db      18773 CCACAGGAAACTGCGCTGATCAGTCCTATCCGTGAATTGACCTATCGCCAACTGTCGACG 18714

Qy       1669 CGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTC 1728
Db      18713 GCGGCGGATCATGTGGCCCGCGCCCTGTTAGCGCTGGGCGTGCAGCATGGCGACCGCGTG 18654

Qy       1729 GCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCA 1788
Db      18653 GCGGTGGTGATGGAAAAAGGCTGGCAGCAGATTGCCGCCGTACACGGCATTTTACGACTG 18594

Qy       1789 GGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTC 1848
Db      18593 GGTGCGGTCTATCTGCCAGTGGATCCGGTGCTACCGCCACAGCGTCGCCAGCTTTTGCTG 18534

Qy       1849 GATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGG 1908
Db      18533 ACGGTGGGCGAGGTGCGGGTACAAGTAACGCAGCCGGGTCTCACGCAATTGGAGCCGTCG 18474

Qy       1909 CCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAG 1968
Db      18473 CTGCCCG------------TGCTGATCATCGACGACGGAATGCTGGACACGCCTGCTGCG 18426

Qy       1969 CCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGA 2028
Db      18425 CCGTTGCCTGAAGTGGCTGGGGATGTCACGGATCTGGCCTATATCATTTTCACTTCCGGC 18366

Qy       2029 TCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTG 2088
Db      18365 TCCACCGGTACCCCGAAAGGAGTGATGATCGACCACCGTGCGGCCATGAACACGCTGGAA 18306

Qy       2089 GACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTG 2148
Db      18305 GACATCAACGAACGCTTTGGCCTCAATGCGCAGGATAGGGTGTTCGGGCTGTCATCATTG 18246

Qy       2149 AGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTG 2208
```

```
Db     18245 AGCTTTGACCTGTCGGTTTACGATGCCTTTGCGCCTTTTATGGTGGGTGCAGCGCTGGTA 18186

Qy      2209 GTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGCAGAGTTGATCGAACGAGAG 2268
Db     18185 CTGCCGGAAGCAGGACGGGAAAAAGATCCGCGTCATTGGCAGACAGTTATGGCACACGGT 18126

Qy      2269 AAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTGAG 2328
Db     18125 CATGTAAGCGTCTGGAATGCAGTGCCCGCACTGATGCAGATGCTGTGCG---AATACCAC 18069

Qy      2329 GGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATC 2388
Db     18068 AGCGGCGATCGGATGAGTTATCCGACGTTGCGTCTGGCACTGTTGAGCGGCGACTGGATC 18009

Qy      2389 CCGGTGGGCCTGCCTGGCGA---GCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGC 2445
Db     18008 CCGCTAACGTTACCGGAGCAGATGCGCGAGCGGCTCAATGAAACGATGGACATCATCAGT 17949

Qy      2446 CTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGAC 2505
Db     17948 CTGGGTGGAGCGACCGAGTGCGCCATCTGGTCGGTCTACTACCCGATAGGTGAGGTGGAA 17889

Qy      2506 CTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTC 2565
Db     17888 TCGACGTGGACCAGTATTCCCTACGGTCGGGGCCTGCGCAACCAGCCAGTATACGTGCTA 17829

Qy      2566 GATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTC 2625
Db     17828 AATGCGCAACTGGAGGAATGTCCGGTCGGGGTGGAAGGAGAGATTTGCATTGGCGGGATG 17769

Qy      2626 GGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCAC 2685
Db     17768 GGGCTGGCACAAGGCTACCTGAACGACGCAGAGAAAACGGCGGCGAGCTTTGTCTGGCGC 17709

Qy      2686 CCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGA 2745
Db     17708 GAAGCGAGTGGTGAGCGAATTTACCGCACTGGGGATCGCGGGCGCTACTTTGCTGACGGG 17649

Qy      2746 AACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAG 2805
Db     17648 CAAGTCGCCTTTTTGGGGCGCAACGATACCCAAGTGAAGGTGAATGGTTACCGTATCGAA 17589

Qy      2806 CTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTG 2865
Db     17588 CTGGGGGAAATCGAGCGCTGCATTGCGCGACATCCCGATGTGGAGCAGTCAGTGGTGGTG 17529

Qy      2866 CCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACA 2925
Db     17528 GCAGTGGGTAA-------------------------------TTCTCAACATCGTCGG 17502

Qy      2926 CGGAGACGCGCTGCCGAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCA 2985
Db     17501 CTGGTCGCTTTTGCCAAACTGCACGATCGCCACCAGGCGCAGGCATTGCAAGCTAAGGAA 17442

Qy      2986 CACGCCGCCGAAGCGGACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGA 3045
Db     17441 ------GCGGAGGCGGCGGCACTGGCGCAGGGTATTATTGTGAATCCGGCACAGCGTCTA 17388

Qy      3046 CACGGACTCCGGAGGGACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCG 3105
Db     17387 GCGTTCAAACTCAAGGAGCCACATATTCGCGCGCTGGATGGTCTGGGCATTGCACTGACG 17328

Qy      3106 CGGGAGGCGGGGCTGGACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCC 3165
Db     17327 GCACCGGCGGATAGCACACGTTACATCAAACGCCGCAGCTATCGTCATTTCAGCGCGCAA 17268

Qy      3166 CCGATTCCGTTTGTTGAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGAC 3225
Db     17267 AAAACCACGCTGGCACAGTTGGGGCAATTGCTGTCGGGCTTGGGGCAGATGCGTCTACCC 17208

Qy      3226 GGCGCGACCCTTCCCAAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACC 3285
Db     17207 GGTCTACCTTTTGCCAAGTATGCCTATGCGTCCGCCGGGGGGCTATACCCGGTGCAAACC 17148

Qy      3286 TACGCGTATGTCAAATCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCAC 3345
Db     17147 TACGTGTACCTGCATCCAGACAAGATCGAAGAGGGAGTATCCGGTATTTACTACTTCGAC 17088

Qy      3346 CCGTTCGAGCACCGTTTGCTGAAGCT---CTCCGATCACGGATCGAGCGCGGAGCGCAC 3402
Db     17087 CCGCGACAGAGCTGTCTTATGCCGGTAGCACCAGAAGTCGAGCTGAACAGTGGTTTTCAT 17028

Qy      3403 GTTCGGCAAAACTTCGACGTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGG 3462
Db     17027 GCCGGACCTAATCAGTCTATTGCCGATCGGGCGGCATTCACGCTGTTTATGGTGGCTGAT 16968

Qy      3463 ATCGACGCCATCGAGTCGCTGTATGG---ATCGTCGTCGCGAGAATTTGCCTGCTGGAG 3519
Db     16967 ATGGCGGTGATCTCGCCATTCTATGGGCAAGAGGCAGCTTGGCACTTCTCGGTGATGGAA 16908

Qy      3520 GCCGGATATATGGCGCAGCTCCTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGT 3579
Db     16907 GCAGGTACTCTCTGCCATTTACTGGAAGAAGATGCGCCGCGCTACGGATTGGGGCTGTGC 16848

Qy      3580 CCGGTGGGGCAATTCAATTTTGAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGAC 3639
Db     16847 CAACTTGGGATGGCAGACTTTTCCGCTGTGGCATCGCATTTTCAATTGTCGCCACATCAT 16788

Qy      3640 GTTTACGTGCACGGCATGCTGGGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACG 3699
Db     16787 CGCTATGTCCATTGCACCGTGGGGGGCGCGATAGGGCAAGAGGCGGCAAGTGCTGCAGCA 16728

Qy      3700 CTCGGTCAGGATTCCTCACCGAGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAG 3759
Db     16727 TTGCTGCGCGATTTCTCCACCTATGAGAAACCGAAGGAAACCGCTGCGCCGCTGGACATG 16668

Qy      3760 CAGCACTTCGCCGATATGCTTCGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTG 3819
Db     16667 CAGAGCTACAAAGATGCCATGCTGCGCGGCCTGCGTCAGCAACTGCCTGACTATATGGTG 16608

Qy      3820 CCTACAGTCTTCGTGGAGCTCGATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGT 3879
Db     16607 CCGAGTGATCTGATGTTAGCGACCGATTTTCCGTTAACCGCTAACGGCAAGCTAGATCGG 16548

Qy      3880 AAGGCCCTGCGCGAGCGGAAGGATACCTCGTCGCCGCGGCAT------TCGGGGCACACG 3933
Db     16547 CAAAAATTACAGCTGCAGGGCGAACAAATTGCCCACCAGCGTGACGGCGTGGGTCCAATC 16488

Qy      3934 GCGCCACGGGACGCCTTGGAGGAGATCCTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTG 3993
Db     16487 CAGGTGGACAGTGCGTTACAACAGCGGCTGGTGGCGCTCTGGCAGGAGGTACTCGGCGTG 16428

Qy      3994 GAGGTGGTCGGGCTCCAGCAGAGCTTCGTCGATCTTGGTGCGACATCGATTCACATCGTT 4053
Db     16427 AGCCACGTGTCGGCCGAAGACGATTTCTTCTCGCTGGGGGGCAGTTCTATAGAATTGGTG 16368

Qy      4054 CGCATGAGGAGCCTGTTGCAGAAGAGGCTGGATAGGGAGATCGCCATCACCGAGTTGTTC 4113
Db     16367 CGTATTCAGCAGGCACTTGAGGCGATTATCGGGCAGGAGATTCCCATTGTCGATCTGTTC 16308

Qy      4114 CAGTACCCGAACCTCGGCTCGCTGGCG 4140
Db     16307 CGTCTGCCAACCATCGCGGATGTGGCG 16281
```

RESULT 5
US-10-085-959-54
; Sequence 54, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870Alel Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648

```
; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE:  2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54
;  LENGTH: 32159
;  TYPE: DNA
;  ORGANISM: Escherichia coli
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (10795)..(10795)
;  OTHER INFORMATION: Unsure
US-10-085-959-54

  Query Match             9.2%;  Score 390.4;  DB 12;  Length 32159;
  Best Local Similarity   51.2%;  Pred. No. 9.4e-92;
  Matches 1086;  Conservative    0;  Mismatches  997;  Indels   38;  Gaps    6;

Qy        761 CGATGGATTACTGGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGA 820
Db      10021 CGCGCGCTTACTGGCTGGCGAAAGCATCGACGCTTCCCCCCGCGCCCGTCTTGCCGCTGG 10080

Qy        821 AGGCCGATCCATCTACCCTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGT 880
Db      10081 CCTGCGAACCCGCCACGCTATGTGAAGTCCGTAATACCCGACGCCGCATGATTGTTCCGG 10140

Qy        881 CGGACTCCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCG 940
Db      10141 CAACACGCTGGCACGCCTTTAGCAACCGGGCCGGCGAGTATGGCGTGACGCCGACAATGG 10200

Qy        941 TCATCCTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTACGC 1000
Db      10201 CGCTGGCGACCTGTTTTTCTGCCGTGCTGGCTCGCTGGGGCGGCCTGACGCGTCTGCTGC 10260

Qy       1001 TCAACATAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGG 1060
Db      10261 TTAACATCACCTTATTCGACCGCCAGCCGCTGCACCCGGCGGTTGGCGCGATGCTTGCCG 10320

Qy       1061 ACTTCACGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGC 1120
Db      10321 ACTTCACCAATATTCTTCTGCTGGACACCGCCTGCGATGGCGATACCGTCAGCAACCTGG 10380

Qy       1121 GCGCTAAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTA 1180
Db      10381 CGCGTAA---AAACCAGCTCACGTTTACGGAGGACTGGGAGCATCGCCACTGGTCCGGCG 10437

Qy       1181 TCGAGGTCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCG 1240
Db      10438 TCGAATTACTCCGTGAACTCAAACGCCAGCAGCGCTACCCCCACGGCGC------CCCGG 10491

Qy       1241 TGGTGCTCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCG 1300
Db      10492 TGGTATTTACCAGCAATCTGGGGCGTTCCCTCTACAGCAGCCGCGCAGAATCGCCGTTGG 10551

Qy       1301 GAACTCCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACG 1360
Db      10552 GCGAGCCGGAATGGGGCATCTCGCAAACGCCGCAGGTCTGGATAGATCATCTGGCGTTCG 10611

Qy       1361 AGCACGATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACC 1420
Db      10612 AGCATCACGGCGAGGTCTGGCTACAATGGGACAGCAACGACGCGCTGTTCCCTCCGGCGT 10671

Qy       1421 TTCTGGACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCAT 1480
Db      10672 TAGTCGAAACATTGTTCGACGCCTACTGCCAGTTGATTAACCAACTCTGCGATGACGAAA 10731

Qy       1481 GGGGTGAACAGGTGC--------GCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGC 1532
Db      10732 GCGCCTGGCAAAAGCCGTTCGCAGATATTATGCCCGCCAGCCAGCGCGCGATACGCGAAC 10791

Qy       1533 GAGCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGC 1592
Db      10792 GGGNTCAACKCCACCGGGTGCCCCCATTCCCGAAGGCTTGCTGCATGAAGGCATTTTCCG 10851

Qy       1593 GCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTA 1652
Db      10852 TATCGCTCTGCAACAGCCGCAGGCGCTGGCGGTAACGGACATGCGTTATCAGTGGAATTA 10911

Qy       1653 CGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACG 1712
Db      10912 TCATGAGCTGACAGACTATGCCCGCCGTTGCGCGGGCAGGTTAATCGAGTGCGGGGTTCA 10971

Qy       1713 CCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCT 1772
Db      10972 GCCCGGCGATAATGTGGCTATCACGATGTCGAAAGGCGCAGGACAACTTGTTGCGGTTCT 11031

Qy       1773 CGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCG 1832
Db      11032 GGCCGTCCTGCTGGCCGGGGCGGTTTACGTTCCGGTTTCGCTGGATCAGCCTGCCGCACG 11091

Qy       1833 TATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGA 1892
Db      11092 GCGCGAGAAAATCTACGCTGACGCCAGCGTCCGGCTGGTGCTCATTTGTCAGCACGACGC 11151

Qy       1893 TGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGA 1952
Db      11152 CAGCGCCGGGTCA-----GACGATATTCCCGCCCTTGCCTGGCAGCAGGCCATT------ 11200

Qy       1953 AGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGT 2012
Db      11201 ----GAGGCGGAGCCGATCGCCAACCCGGTAGTACGCGCCCCCACGCAACCGGCCTACAT 11256

Qy       2013 CATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGC 2072
Db      11257 TATCTACACCTCCGGCTCTACCGGTACGCCGAAAGGGGTAGTCATTTCTCACCGGGGAGC 11316

Qy       2073 CGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCT 2132
Db      11317 GCTTAACACCTGTTGCGATATCAATACCCGCTATCAGGTTGGCCCGCATGACAGGGTGCT 11376

Qy       2133 GGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGC 2192
Db      11377 GGCCCTCTCCGCCCTACATTTTGATTTATCGGTTTACGACATTTTTGGCGTACTGCGCGC 11436

Qy       2193 GGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGA 2252
Db      11437 GGGCGGCGCGCTGGTGATGGTGATGGAAAATCAACGGCGCGATCCTCACGCATGGTGTGA 11496

Qy       2253 GTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCT 2312
Db      11497 GCTGATCCAGCGCCATCAGGTCACGCTCTGGAACAGCGTCCCGGCGCTGTTCGATATGCT 11556

Qy       2313 CGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCT 2372
Db      11557 GCTGACCTGGTGTGAAGGTTTCGCCGACGCCACGCCGGAAAACCTGCGCGCAGTGATGCT 11616

Qy       2373 GAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGT 2432
Db      11617 TTCCGGCGACTGGATCGGGCTTGACCTCCCCGCCCGTTATCGGGCCTTCCGGCCACAAGG 11676

Qy       2433 GTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGT 2492
Db      11677 ACAATTTATCGCGATGGGCGGCGCCACCGAGGCGTCTATCTGGTCTAACGCCTGCGAAAT 11736

Qy       2493 GAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGAC 2552
Db      11737 TCACGACGTCCCCGCCCACTGGCGCTCCATCCCTTACGGTTTTCCGCTAACCAACCAACG 11796

Qy       2553 GTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTA 2612
Db      11797 CTACCGGGTGGTGGATGAACAGGGCCGGGACTGCCCTGACTGGGTGCCGGGTGAATTATG 11856

Qy       2613 CATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAG 2672
```

```
Db      11857 GATTGGCGGCATTGGGGTCGCGGAAGGCTATTTCAACGATCCCCTGCGTAGCGAGCAGCA 11916

Qy       2673 CTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTA 2732

Db      11917 ATTTTTGACGCTCCCGGA------CGAGCGCTGGTATCGCACCGGCGATCTCGGCTGCTA 11970

Qy       2733 CCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGG 2792

Db      11971 CTGGCCAGATGGCACAATCGAGTTCCTCGGTCGTCGCGACAAGCAGGTCAAAGTCGGAGG 12030

Qy       2793 ATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGA 2852

Db      12031 ATATCGCATCGAGCTGGGCGAAATCGAAAGCGCGCTCAGCCAGCTGGCGGGGGTGAAACA 12090

Qy       2853 CGCGGTGATTGTGCCCGTCGG 2873

Db      12091 AGCAACCGTTCTGGCGATCGG 12111
```

RESULT 6
US-10-314-657-1
; Sequence 1, Application US/10314657
; Publication No. US20030175888A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: CHENG, Yi-Qiang
; APPLICANT: TANG, Gong-Li
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; TITLE OF INVENTION: Synthases and Methods of Use
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1

```
  Query Match             8.5%;  Score 360.6;  DB 12;  Length 135638;
  Best Local Similarity   55.4%;  Pred. No. 7.6e-84;
  Matches  768;  Conservative    0;  Mismatches  604;  Indels   15;  Gaps    3;

Qy       1541 ACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCG 1600

Db      72583 ACACCACCCGCGCGTTCCCCGACGACCGGCTGATGCACCAGCTCTTCGAGGAGCAGGCGG 72642

Qy       1601 AGCAGCTGCCCATGCAGCTCGCCGTGGTGTCG---GCGCGCAAGACGCTCACGTACGAAG 1657

Db      72643 CCCAGCAGCCTCGGGCGGAGGCGCTGCGCTGGCGCGGCGGCGGCACGATGACATACCAGG 72702

Qy       1658 AGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGA 1717

Db      72703 AGCTGAACCGGCGGGCCAACCGCATCGCCGCCCGGCTGGCCGCCGAGGACGTGGGCCCGG 72762

Qy       1718 ACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGG 1777

Db      72763 AGACGGTCGTCGCCGTCAGTGTGCCGCGCGGGCCCATGATGGTCGCGGTGGTCCTCGGGA 72822

Qy       1778 TGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCC 1837

Db      72823 TCCTGAAGGCCGGCGGTGTCTATCTGCCGATGGAGCCGCATCTGCCCGCCGAGCGGGCCG 72882

Qy       1838 ACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCA 1897

Db      72883 CCGTCATCCTCGAGGAGGCGCACGCCGAGGTCGTCGTCACCACCGCCGACCGCGAGGGCT 72942

Qy       1898 AAC---TGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAG 1954

Db      72943 GGCCGGTGCCCGACGGATACGCCCGCGTGTGCGCGGACGCCGCCGTGGAGGGCCCCCACC 73002

Qy       1955 GCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCA 2014

Db      73003 CCGCCGACGCGGACAACTGCCCGCGGCCCGTGACGCAGCCGCACAACACGGCGTACATCA 73062

Qy       2015 TCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCG 2074

Db      73063 TCTTCACCTCGGGCAGCACCGGCCGCCCCAAGGGTGTGGCGGTCGCGCACCGGCCCGTCC 73122

Qy       2075 TCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGG 2134

Db      73123 TGAACCTCATCAACTGGTGCCGGCGCACCTTCGGGTTCGGGCCCGGGGACATGGGCCTGT 73182

Qy       2135 CGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGG 2194

Db      73183 GTGTCACCTCGCTCGGCTTCGACCTGTCCGTGTTCGACGTGTTCGGCCTGCTCGGTACGG 73242

Qy       2195 GCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGT 2254

Db      73243 GTGCCGCGCTCTACATCGCGGACGCGGAGCAGCAGCGCGATCCCGCTCTCCTCCTGGACG 73302

Qy       2255 TGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCG 2314

Db      73303 TGCTGATCGAGGAGCCGGTCACCTTCTGGAACTCGGCGCCCACCACGCTCGCCCAGGTAG 73362

Qy       2315 TCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGA 2374

Db      73363 GTCCGCTGCTCGACACAGTGGGCACGGCGGGTACCGGTGATCTGCGCCTGGTGTTCCTCA 73422

Qy       2375 GCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGT 2434

Db      73423 GCGGCGACTTCACGCCGCTGCCCCTCCCCGACGAGGTGCGCGCCGTCTTCCCGCGCGCCG 73482

Qy       2435 CGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGA 2494

Db      73483 ACATGATCAGCCTGGGCGGTGCCACCGAGGCGACGGTGTGGTCCAACTGGTTCCGCATCG 73542

Qy       2495 GGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGT 2554

Db      73543 GGGCGATCGACCCGGCCTGGCGCAGCATCCCCTACGGCAGGCCCATCGACAACAGCCGCT 73602

Qy       2555 TCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACA 2614

Db      73603 ACCACGTGCTCGACGAGGCGCTGCGCCCCTGTCCCGTCGGCGTCGAGGGGGATCTGTACA 73662

Qy       2615 TTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCT 2674

Db      73663 TCGGCGGTGAGTGTCTGGCCCTCGGCTACGTCAACCAGCCCGAGCTGACCGCCGACCGGT 73722

Qy       2675 TCCTC---------GTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGG 2725

Db      73723 TCATCCCCGACCCGTTCCACGAGGACCCTCAGGAGCGCCTCTACAAGACCGGTGACAGGG 73782

Qy       2726 GCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGC 2785

Db      73783 CCCTGTACTACCCGGACGGCAACCTGAGCTTCCAGGGGCGGGCCGACGGACAGGTGAAGG 73842

Qy       2786 TTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACG 2845

Db      73843 TCCGCGGGTTCCGGGTGGAACTGGCCGAGATCGAGCACCGGCTGCGGGCGCACGACGGGG 73902

Qy       2846 TACGCGACGCGGTGATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCT 2905

Db      73903 TCAAGGACGCCGTGGTCCTGGCCCGGGAGGACGGCTGCGGGGACCGCACCCTCGTCGCGT 73962

Qy       2906 ATGTGGT 2912

Db      73963 ACCTCGT 73969
```

RESULT 7

```
US-09-815-242-7601
; Sequence 7601, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7601
;   LENGTH: 1416
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1416)
US-09-815-242-7601

  Query Match             7.2%;  Score 305.6;  DB 9;  Length 1416;
  Best Local Similarity   54.4%;  Pred. No. 7.6e-70;
  Matches  692;  Conservative    0;  Mismatches  559;  Indels   21;  Gaps    3;

Qy       1602 GCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCT 1661
Db         57 GCAACAGCCGCAGGCGCTGGCGGTAACGGACATGCGTTATCAGTGGAATTATCATGAGCT 116

Qy       1662 TTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACAC 1721
Db        117 GACAGACTATGCCCGCCGTTGCGCGGGCAGGTTAATCGAGTGCGGGGTTCAGCCCGGCGA 176

Qy       1722 ATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCT 1781
Db        177 TAATGTGGCTATCACGATGTCGAAAGGCGCAGGACAACTTGTTGCGGTTCTGGCCGTCCT 236

Qy       1782 CGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTA 1841
Db        237 GCTGGCCGGGGCGGTTTACGTTCCGGTTTCGCTGGATCAGCCTGCCGCACGGCGCGAGAA 296

Qy       1842 CCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACT 1901
Db        297 AATCTACGCTGACGCCAGCGTCCGGCTGGTGCTCA------TTTGTCAGCACGACGCCAG 350

Qy       1902 GTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGG 1961
Db        351 CGCCGGGTCAGACGATATTCCCGTCCTTGCCTGGCAGCAGGCCATTGAGGCGGAGCCGAT 410

Qy       1962 CGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACAC 2021
Db        411 CGCCAACCC---------GGTAGTACGCGCCCCCACGCAACCGGCCTACATTATCTACAC 461

Qy       2022 CTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACAC 2081
Db        462 CTCCGGCTCTACCGGTACGCCGAAAGGGGTAGTCATTTCTCACCGGGGAGCGCTTAACAC 521

Qy       2082 CATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTC 2141
Db        522 CTGTTGCGATATCAATACCCGCTATCAGGTTGGCCCGCATGACAGGGTGCTGGCCCTCTC 581

Qy       2142 CTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTAC 2201
Db        582 CGCCCTACATTTTGATTTATCGGTTTACGACATTTTTGGCGTACTGCGCGCGGGCGGCAC 641

Qy       2202 GATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGA 2261
Db        642 GCTGGTGATGGTGATGGAAAATCAACGGCGCGATCCTCACGCATGGTGTGAGCTGATCCA 701

Qy       2262 ACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCA 2321
Db        702 GCGCCATCAGGTCACGCTCTGGAACAGCGTCCCGGCGCTGTTCGATATGCTGCTGACCTG 761

Qy       2322 TTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGA 2381
Db        762 GTGTGAAGGTTTCGCCGACGCCACGCCGGAAAACCTGCGCGCAGTGATGCTTTCCGGCGA 821

Qy       2382 CTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGAT 2441
Db        822 CTGGATCGGGCTTGACCTCCCCGCCCGTTATCGGGCCTTCCGGCCACAAGGACAATTTAT 881

Qy       2442 CAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGT 2501
Db        882 CGCGATGGGCGGCGCCACCGAGGCGTCTATCTGGTCTAACGCCTGCGAAATTCACGACGT 941

Qy       2502 CGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGT 2561
Db        942 CCCCGCCCACTGGCGCTCCATCCCTTACGGTTTTCCGCTAACCAACCAACGCTACCGGGT 1001

Qy       2562 GCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGG 2621
Db       1002 GGTGGATGAACAGGGCCGGGACTGCCCTGACTGGGTGCCGGGTGAATTATGGATTGGCGG 1061

Qy       2622 GGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGT 2681
Db       1062 CATTGGGGTCGCGGAAGGCTATTTCAACGATCCCCTGCGTAGCGAGCAGCAATTTTTGAC 1121

Qy       2682 GCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGA 2741
Db       1122 GCTCCCGGA------CGAGCGCTGGTATCGCACCGGCGATCTCGGCTGCTACTGGCCAGA 1175

Qy       2742 TGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGT 2801
Db       1176 TGGCACAATCGAGTTCCTCGGTCGTCGCGACAAGCAGGTCAAAGTCGGAGGATATCGCAT 1235

Qy       2802 TGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGAT 2861
Db       1236 CGAGCTGGGCGAAATCGAAAGCGCGCTCAGCCAGCTGGCGGGGGTGAAACAAGCAACCGT 1295

Qy       2862 TGTGCCCGTCGG 2873
Db       1296 TCTGGCGATCGG 1307


RESULT 8
US-10-156-761-3143
; Sequence 3143, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
```

```
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3143
;   LENGTH: 1875
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1875)
US-10-156-761-3143

  Query Match             7.0%;  Score 296.8;  DB 14;  Length 1875;
  Best Local Similarity   53.0%;  Pred. No. 1.6e-67;
  Matches  743;  Conservative    0;  Mismatches  632;  Indels   27;  Gaps    4;

Qy       1533 GAGCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGC 1592
Db         72 GAACGACACCGCGCGGGAGCTGCCCGGCGGTGGCGACACCCTGCCCGAGCTGTTCCGTAC 131

Qy       1593 GCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTA 1652
Db        132 GCAGGTGTCCCGGACCCCTGACGCCACCGCCCTGGTCTTCCGGGACACCCGCCTCAGCTA 191

Qy       1653 CGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACG 1712
Db        192 CACCGAGCTGGACGAGCGCTCGGACCGCCTGGCCCGGCTGCTGGCCGGGCACGGGGCCGC 251

Qy       1713 CCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCT 1772
Db        252 CCCGGACCGGGTCGTGGCCCTGACCGTGCCGCGCTCGGTGGAACTCGTCGTCGCGCTGCT 311

Qy       1773 CGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCG 1832
Db        312 CGCCGTACTCAAGACGGGCGCCGCCTACCTCCCGGTGGACCCGGACTACCCGGCGGACCG 371

Qy       1833 TATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGA 1892
Db        372 CATCGGCTACCTCCTCGCCGACGCCGCCCCGGTCCTGCTGCTCACCCACTCCGACGTGGA 431

Qy       1893 TGGCAAACTGT---------CATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGC 1943
Db        432 GGCGACGCTGCCGCCGGGCGCACGGCCCGAGGGGCTGCCCCGTCTGGTCGTCGACGCACA 491

Qy       1944 CGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCT 2003
Db        492 GGAAGAGCCGCTGGAGCGCGCCCCCTTCGCGCCCGCCCGTCCGCCCGCCGCGCAGGACGC 551

Qy       2004 CGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCA 2063
Db        552 CGCGTACGTCATCTACACCTCCGGCTCCACCGGGCGGCCCAAGGGCGTGGTCGTCCCGCA 611

Qy       2064 TCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGA 2123
Db        612 CGCCGGCATAGTCAACCGGCTGCACTGGATGCAGCACGAGTACCGGCTGACCGGCGACGA 671

Qy       2124 CAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGAT 2183
Db        672 CCGCGTCCTGCAGAAGACGCCCTCGGGCTTCGACGTCTCGGTGTGGAGTTCTTCTGGCC 731

Qy       2184 CCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCA 2243
Db        732 GCTGATCACCGGCGCCACCCTGGTCGTCGCCGAACCCGGCGGGCACAAGGACCCCGCCTA 791

Qy       2244 TTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGAT 2303
Db        792 CCTGGCCGAGCTGATCCGGGGCGAGCGGATCACCACCGTGCACTTCGTCCCGTCGATGCT 851

Qy       2304 GCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCT 2363
Db        852 CCAGGTGTTC---------CTGGACGAGCCGGCCGCGGCGCTCTGCACCGGCCTGCGCCG 902

Qy       2364 TTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAG 2423
Db        903 GGTGGTGTGCAGCGGTGAGGCGCTGCCGGTCGAACTCGCCGACCGCTTCACGGCGTCGCT 962

Qy       2424 GCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGG 2483
Db        963 GCCCGGCGTGCCCCTGCACAACCTGTACGGCCCGACCGAGGCATCGGTC---GACGTCAC 1019

Qy       2484 GTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCG 2543
Db       1020 GTACTGGGAGTACACCCCCGAGCCCGGCGCCACCTCGGTGCCGATCGGCCGGCCCGTGTG 1079

Qy       2544 CAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGG 2603
Db       1080 GAACACCCGCCTGTACGTCCTGGACGCGGCGCTGCGCCCGGCCGCGGTGGGGGAGACCGG 1139

Qy       2604 GCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGAC 2663
Db       1140 CGAGCTGTATCTGGCCGGGGTGCAGCTGGCGCTCGGCTATCTGAACCGCCCCGGCCTGAC 1199

Qy       2664 GCGCAAGAGCTTCCTCGTGCACCCCGA------GACCGGGGAGCGCCTCTACAAGACCGG 2717
Db       1200 GGCGGAACGCTTCGTCGCCGACCCGTACGGGCCGCCGGGCAGCCGTATGTACCGCACCGG 1259

Qy       2718 CGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCA 2777
Db       1260 GGACCTGGCCCGCCGCCGCCCGGACGGCGCGCTGGACTACGTGGGACGCGCGGACCACCA 1319

Qy       2778 AATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCA 2837
Db       1320 GGTCAAGATCCGGGGCCTGCGCGTCGAACTGGGCGAGATCGAGGCGGCCCTGGCCCGGCA 1379

Qy       2838 TCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCT 2897
Db       1380 TCCGGCGGTCTCCGGGGCGGCGGTGCTGGTCCGGGAGGACCGGCCGGGCGACAAGCGGCT 1439

Qy       2898 TCTAGCCTATGTGGTCCCGGAG 2919
Db       1440 GGTGGGCTATGTGGTCCCGGCG 1461

RESULT 9
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

  Query Match             7.0%;  Score 296.8;  DB 14;  Length 9025608;
  Best Local Similarity   53.0%;  Pred. No. 7.8e-67;
  Matches  743;  Conservative    0;  Mismatches  632;  Indels   27;  Gaps    4;

Qy       1533 GAGCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGC 1592
Db    3932113 GAACGACACCGCGCGGGAGCTGCCCGGCGGTGGCGACACCCTGCCCGAGCTGTTCCGTAC 3932054

Qy       1593 GCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTA 1652
Db    3932053 GCAGGTGTCCCGGACCCCTGACGCCACCGCCCTGGTCTTCCGGGACACCCGCCTCAGCTA 3931994

Qy       1653 CGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACG 1712
Db    3931993 CACCGAGCTGGACGAGCGCTCGGACCGCCTGGCCCGGCTGCTGGCCGGGCACGGGGCCGC 3931934

Qy       1713 CCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCT 1772
Db    3931933 CCCGGACCGGGTCGTGGCCCTGACCGTGCCGCGCTCGGTGGAACTCGTCGTCGCGCTGCT 3931874

Qy       1773 CGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCG 1832
Db    3931873 CGCCGTACTCAAGACGGGCGCCGCCTACCTCCCGGTGGACCCGGACTACCCGGCGGACCG 3931814

Qy       1833 TATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGA 1892
Db    3931813 CATCGGCTACCTCCTCGCCGACGCCGCCCCGGTCCTGCTGCTCACCCACTCCGACGTGGA 3931754

Qy       1893 TGGCAAACTGT---------CATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGC 1943
Db    3931753 GGCGACGCTGCCGCCGGGCGCACGGCCCGAGGGGCTGCCCCGTCTGGTCGTCGACGCACA 3931694

Qy       1944 CGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCT 2003
Db    3931693 GGAAGAGCCGCTGGAGCGCGCCCCCTTCGCGCCCGCCCGTCCGCCCGCCGCGCAGGACGC 3931634

Qy       2004 CGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCA 2063
Db    3931633 CGCGTACGTCATCTACACCTCCGGCTCCACCGGGCGGCCCAAGGGCGTGGTCGTCCCGCA 3931574

Qy       2064 TCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGA 2123
Db    3931573 CGCCGGCATAGTCAACCGGCTGCACTGGATGCAGCACGAGTACCGGCTGACCGGCGACGA 3931514

Qy       2124 CAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGAT 2183
Db    3931513 CCGCGTCCTGCAGAAGACGCCCTCGGGCTTCGACGTCTCGGTGTGGGAGTTCTTCTGGCC 3931454

Qy       2184 CCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCA 2243
Db    3931453 GCTGATCACCGGCGCCACCCTGGTCGTCGCCGAACCCGGCGGGCACAAGGACCCCGCCTA 3931394

Qy       2244 TTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGAT 2303
Db    3931393 CCTGGCCGAGCTGATCCGGGGCGAGCGGATCACCACCGTGCACTTCGTCCCGTCGATGCT 3931334

Qy       2304 GCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCT 2363
Db    3931333 CCAGGTGTTC---------CTGGACGAGCCGGCCGCGGCGCTCTGCACCGGCCTGCGCCG 3931283

Qy       2364 TTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAG 2423
Db    3931282 GGTGGTGTGCAGCGGTGAGGCGCTGCCGGTCGAACTCGCCGACCGCTTCACGGCGTCGCT 3931223

Qy       2424 GCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGG 2483
Db    3931222 GCCCGGCGTGCCCCTGCACAACCTGTACGGCCCGACCGAGGCATCGGTC---GACGTCAC 3931166

Qy       2484 GTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCG 2543
Db    3931165 GTACTGGGAGTACACCCCCGAGCCCGGCGCCACCTCGGTGCCGATCGGCCGGCCCGTGTG 3931106

Qy       2544 CAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGG 2603
Db    3931105 GAACACCCGCCTGTACGTCCTGGACGCGGCGCTGCGCCCGGCCGCGGTGGGGAGACCGG 3931046

Qy       2604 GCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGAC 2663
Db    3931045 CGAGCTGTATCTGGCCGGGGTGCAGCTGGCGCTCGGCTATCTGAACCGCCCCGGCCTGAC 3930986

Qy       2664 GCGCAAGAGCTTCCTCGTGCACCCCGA------GACCGGGGAGCGCCTCTACAAGACCGG 2717
Db    3930985 GGCGGAACGCTTCGTCGCCGACCCGTACGGGCCGCCGGGCAGCCGTATGTACCGCACCGG 3930926

Qy       2718 CGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCA 2777
Db    3930925 GGACCTGGCCCGCCGCCGCCCGGACGGCGCGCTGGACTACGTGGGACGCGCGGACCACCA 3930866

Qy       2778 AATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCA 2837
Db    3930865 GGTCAAGATCCGGGGCCTGCGCGTCGAACTGGGCGAGATCGAGGCGGCCCTGGCCCGGCA 3930806

Qy       2838 TCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCT 2897
Db    3930805 TCCGGCGGTCTCCGGGGCGGCGGTGCTGGTCCGGGAGGACCGGCCGGGCGACAAGCGGCT 3930746

Qy       2898 TCTAGCCTATGTGGTCCCGGAG 2919
Db    3930745 GGTGGGCTATGTGGTCCCGGCG 3930724


RESULT 10
US-10-156-761-3184
; Sequence 3184, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3184
;   LENGTH: 5304
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(5304)
US-10-156-761-3184

  Query Match             6.4%;  Score 270.4;  DB 14;  Length 5304;
  Best Local Similarity   52.0%;  Pred. No. 1.5e-60;
  Matches  753;  Conservative    0;  Mismatches  656;  Indels   39;  Gaps    5;

Qy       1503 GCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTGAG 1562
Db        741 GCGAGCGGACGCCGACCGGGCCGTGCTCGCGCGGGCCAACGGCACCGACGCCGCGTTCGA 800

Qy       1563 CGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCA---GCT 1619
```

```
Db     801 GGACACCGTGCGCCTCGACGAGCTGTTCACCCGGCAGGCGCGGCTGACCCCGGACGCCAC 860

Qy    1620 CGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCG 1679
Db     861 CGCCGTGGTCGGCGACGGCTTCCGGCTCACCTACGCCGAACTCGACGCCCGTGCCAACCG 920

Qy    1680 ACTTGGCGCGCGGCTGCGCGAGCAGGGGCACGCCCGAACACATTGGTCGCGGTGGTGAT 1739
Db     921 GCTGGCGCGGACGCTGCGCGAGCGGGGCGTGGGCCGCGACGACCTCGTCGCGGTGCTGGC 980

Qy    1740 GGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTA 1799
Db     981 CGACCGCTCGCCGCAGCTGCTCGTCGCCGTCTACGCCGTCCTCAAGGCCGGCGCCGCCTA 1040

Qy    1800 CGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGA 1859
Db    1041 TCTGCCGGTGGACCCCGGCTACCCCGTCGAGCGCATCCGCTACGTCCTGGACGACAGCGC 1100

Qy    1860 GGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGAT 1919
Db    1101 CGCTCGTCTCGTCCTCAGCACGCCCCGCCACCGCGACCT---------GGTCCACGAGGG 1151

Qy    1920 CCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGAT 1979
Db    1152 CCGCGAGGTACTCGACCTGACCGACGAGTCCGCGTACGCAAAGGACAGCACGGCGCCGGA 1211

Qy    1980 GCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTT 2039
Db    1212 GGTCATCGGCGACGCCCGCGACCTGGCGTACGTCATCTACACCTCGGGCTCGACCGGCAA 1271

Qy    2040 GCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGA 2099
Db    1272 TCCCAAGGGCGTCCAGATCGAGCACCGCTCGGTGGTCAACCGCATCAACTGGATGCAGCG 1331

Qy    2100 GCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCT 2159
Db    1332 CGCCTACCCCGTCACGTCCGCCGACGTGATCATGCAGAAGACGCCCACCTCCTTCGACGT 1391

Qy    2160 CTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGC 2219
Db    1392 GTCCGTCTGGGAGCTGTTCTGGTGGATGTTCGAGGGCGCCGCGGTGTGCCTGCCCGACCC 1451

Qy    2220 GTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGT 2279
Db    1452 CGGCGCGGAACGCGACCCGGAGGCGCTGGTCACCGCCATCGAGCGGTACGGCGTGACCAC 1511

Qy    2280 GTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGA 2339
Db    1512 CATGCACTTCGTGCCGTCCATGCTGGGCGCCTTCCTCGACTTCGTCGTGGGAGCGGGCGT 1571

Qy    2340 TTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCT 2399
Db    1572 ACAGGCCGGGCTGGTGTCGCTGCGCCAGGTCTTCGCCAGCGGTGAGGCGCTCACGCCGCA 1631

Qy    2400 GCCTGGCGAGCTCCAGG-----------------CCATCAGGCCCGGCGTGTCGGTGAT 2441
Db    1632 CCACGTCCGGCGCTTCGCCGAGGTCGCCGCCGCCACCACGGGCCCGGCCGCCCGGCTGAT 1691

Qy    2442 CAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGT 2501
Db    1692 CAACCTGTACGGCCCCACCGAGGCCACCGTCGACGTGACCCACCACGCGTGCGGGCCCGA 1751

Qy    2502 CGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGT 2561
Db    1752 CGACGT------GGACCGGGTCCCGATCGGCCGCCCGATCGACAACACCCGCATCCTCGT 1805

Qy    2562 GCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGG 2621
Db    1806 CCTCGACTCCGGACTGCGGCCCCAGCCGGTCGGTGTGCCCGGCGAACTGTGCATCGCCGG 1865

Qy    2622 GGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGT 2681
Db    1866 TGTCGGTCTGGCCCGCGGCTACCTCGGCCGGCCGGAACTGACCGCCGAGCGGTTCGTCGC 1925

Qy    2682 GCACCCCGAGACCGG---GGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCC 2738
Db    1926 CGCCCCGGCCGCCGGTGAGGAGCGCGTCTACCGCACGGGCGACCTGGCCCGGTGGCTGCC 1985

Qy    2739 CGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCG 2798
Db    1986 CGACGGCAGCATCGACTACCTGGGCCGTATCGACCACCAGGTCAAGGTGCGCGGCTACCG 2045

Qy    2799 CGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGT 2858
Db    2046 CATCGAGCTCGGCGAGATCGAGGAGTCGCTGCGCCGCCACCCCCAGGTCAAGGACGCCGT 2105

Qy    2859 GATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGA 2918
Db    2106 GGTGCTCGCCCGCAAGGCCGAGGACGGTCAGACCCACCTGTGCGGCTACGTCCAGTCGGC 2165

Qy    2919 GGGCACAC 2926
Db    2166 GGACGCCC 2173
```

RESULT 11
US-09-815-242-4052
; Sequence 4052, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4052
; LENGTH: 13029
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4052

Query Match 6.2%; Score 261.4; DB 9; Length 13029;
Best Local Similarity 51.8%; Pred. No. 4.1e-58;
Matches 704; Conservative 0; Mismatches 631; Indels 24; Gaps 4;

```
Qy    1572 GCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTC 1631
Db    3387 GCTGCCGGAGCTGCTGGAGCGGCAACTGGCGCAGTCCGCCGAACGCGTCGCCCTGGAGTG 3446

Qy    1632 GGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCG 1691
Db    3447 GGACGGCGGCAGCCTCGGCTACGCCGAACTGCACGCCCGGGCCAACCGCCTGGCCCACTA 3506
```

```
Qy       1692 GCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTG 1751
Db       3507 CCTGCGCGACAAGGGCGTCGGCCCGGACGTGCGGGTAGCGATCTGCGCCGAGCGTTCGCC 3566

Qy       1752 GGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGA 1811
Db       3567 GCAACTGCTGGTCGGTCTGCTGGCGATCGTCAAGGCGGGCGGTGCCTACGTGCCGCTGGA 3626

Qy       1812 TGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGT 1871
Db       3627 CCCGGACTACCCCAGCGAACGCCTGGCCTACATGCTCGCCGACAGTGGCGTCGAGCTGCT 3686

Qy       1872 GCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCT 1931
Db       3687 ACTGACCCAGGCCCATCTGTTCGAACGCCTGCCTGGCGCGGAGGGCGTCACGCCGATCTG 3746

Qy       1932 CGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGAC 1991
Db       3747 CCTGGACAGCCTCAAGCTCGACAACTGGCCGAGCCAGGCACCGGGCCTGCACCTGCACGG 3806

Qy       1992 ACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGT 2051
Db       3807 CGACA---ATCTCGCCTACGTCATCTACACCTCCGGCTCCACCGGCCAGCCGAAGGGCGT 3863

Qy       2052 GATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAAT 2111
Db       3864 GGGCAACACCCACGCGGCGCTGGCCGAGCGCCTGCAATGGATGCAGGCCACCTACACCCT 3923

Qy       2112 AGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGA 2171
Db       3924 CGACGGCGACGACGTGCTGATGCAGAAGGCGCCGGTGAGCTTCGACGTGTCGGTGTGGGA 3983

Qy       2172 TGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCG 2231
Db       3984 GTGCTTCTGGCCGCTGGTCACCGGTTGCCGGTTGGTCCTCGCCGCCCCCGGCGAGCACCG 4043

Qy       2232 CGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGT 2291
Db       4044 CGATCCGGCGCGCCTGGTGGAACTGGTGCGGCAGTTCGGGGTGACCACGCTGCACTTCGT 4103

Qy       2292 GCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAG 2351
Db       4104 GCCGCCGCTGCTGCAACTGTTCATC---------GACGAACCCGGCGTCGCCGCCTGCGG 4154

Qy       2352 GTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCT 2411
Db       4155 CAGCCTGCGCCGCCTGTTCTCCGGCGGCGAGGCGCTTCCGGCGGAGCTGCGCAACCGCGT 4214

Qy       2412 CCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGAT 2471
Db       4215 GTTGCAACGCCTGCCGGCGGTGGCCCTGCATAACCGCTACGGGCCGACCGAGACCGCCAT 4274

Qy       2472 CTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGG 2531
Db       4275 ------CAACGTCACCCATTGGCAGTGCCGCGCGGAAGACGGCGAGCGTTCGCCGATCGG 4328

Qy       2532 CCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGT 2591
Db       4329 CCGACCGCTGGGCAACGTGGTCTGCCGCGTGCTGGACGCCGAGTTCAACCTGCTGCCGGC 4388

Qy       2592 CTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGA 2651
Db       4389 CGGCGTCGCCGGCGAGCTGTGCATCGGCGGCCTGGGCCTGGCCCGCGGCTACCTGGGGCG 4448

Qy       2652 TGAAGAGAAGACGCGCAAGAGCTTCCTCG------TGCACCCCGAGACCGGGGAGCGCCT 2705
Db       4449 CCCGGCGCTGAGCGCCGAGCGCTTCGTCGCCGATCCGTTCTCGGCCGCCGGCGAGCGCCT 4508

Qy       2706 CTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCG 2765
Db       4509 GTACCGCACCGGCGACCGCGCGCGCTGGAACGCCGACGGCGTGCTGGAATACCTCGGCCG 4568

Qy       2766 TGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAAC 2825
Db       4569 CCTCGACCAGCAGGTCAAGCTGCGTGGCTTCCGCATCGAGCCGGAGGAAATCCAGGCACG 4628

Qy       2826 GCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGACGCGGC 2885
Db       4629 CCTGCTGGCGCAGCCGGGGGTGGCCCAGGCGGTAGTGGTGATCCGCGAAGGCGTGGCCGG 4688

Qy       2886 GAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCAC 2924
Db       4689 CAGCCAGTTGGTCGGCTACTACACCGGCGCTGTCGGCGC 4727
```

RESULT 12
US-10-156-761-3629
; Sequence 3629, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3629
;   LENGTH: 11058
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(11058)
US-10-156-761-3629

```
  Query Match             5.8%;  Score 246.2;  DB 14;  Length 11058;
  Best Local Similarity   51.5%;  Pred. No. 3.7e-54;
  Matches  709;  Conservative    0;  Mismatches  638;  Indels   30;  Gaps    5;

Qy       1574 TGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGG 1633
Db       1469 TGCACGAGGTGTTCGAGGAGCAGGTGGCGCGGACCCCCGAGGGGGTCGCGGTCTCCTTCG 1528

Qy       1634 CGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGC 1693
Db       1529 AGGGGACGGATCTGACGTACGCGGAGCTGAACGCCCGCGCCAACCGGATCGCCCACGCCC 1588

Qy       1694 TGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGG 1753
Db       1589 TGATCGCGAAGGGGGTCGGCCCGGAGACGCTCGTCGGGCTGAGCCTGGAGCGCGGCATCG 1648

Qy       1754 AGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATG 1813
Db       1649 AACTGATCCCGGCCCTGCTCGGCATCCTCAAGTCGGGTGCCGCGTATCTGCCGCTGGACC 1708

Qy       1814 CCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGC 1873
Db       1709 CGGCGAACCCGGCCGACCGGATCGCGTACATCGTCGGGGACGCACAGGCCCCGGTGGTGG 1768

Qy       1874 TGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCG 1933
Db       1769 TGAC--------CACCGCCGAGCACGCGCATCTGTTCGACGTGGA-GCTGCTGCTCCTCG 1819

Qy       1934 TGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACAC 1993
```

```
Db       1820 ACACGGACGCGCCGGACCTGGCCGCGCGGCCGGACTCCGACCCGGGCGTGCCCGGCAGCC 1879

Qy       1994 CTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGA 2053

Db       1880 CCGAGAACCTGATCTACACGATCTACACCTCCGGCTCGACGGGCCGGCCGAAGGGCGTCG 1939

Qy       2054 TGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAG 2113

Db       1940 CCCTCACCCACACCAACGTCGTACGCCTCCTGGAGCGCGGCAACGAGCACTACGCCTTCA 1999

Qy       2114 GGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATG 2173

Db       2000 CCGACACCGACGTCTGGCCCCTCTTCCACAGCTACGCCTTCGACGTCTCCGTCTGGGAGA 2059

Qy       2174 TGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCG 2233

Db       2060 TGTGGGGTGCCCTGCTGCACGGCGGCCGGCTGATCGTCGTCCCCTACGACGTCACCCGCT 2119

Qy       2234 ATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGC 2293

Db       2120 CCCCGGAGGAGTTCCTCGACCTGCTGGTGCGCGAGCGCGTCACCGTGCTCAACCAGACGC 2179

Qy       2294 CGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGT 2353

Db       2180 CCTCCGCCTTCCGTTCGCTGGTCGCCGCGGACGCCGAACTGTCGCTGCGCGCGGTGGTCT 2239

Qy       2354 CTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCC 2413

Db       2240 TCGCGGGCGAGAAGCTGGAGATCTCCGAACTGCGCCCCTGGGCCGACCGGTTCGGGCTCG 2299

Qy       2414 AGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCT 2473

Db       2300 ACCGGATCGCGCTCG------------TCAACATGTACGGGATCACCGAGACCACCGTCC 2347

Qy       2474 GGTCCATCGGGTACCCCGTGAGGAACGTCGACCT---ATCGTGGGCGAGCATCCCCTACG 2530

Db       2348 ACACCACCTACCACCGGCTGACCGAGCGGGACCTGGACCCGCGGGCGGGCAACGCGGTCG 2407

Qy       2531 GCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGG 2590

Db       2408 GCCATCCGCTGGCCGACCTGCGGGTGTACCTCCTGGACGCGGACCGACAGCTGGTCCCGG 2467

Qy       2591 TCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCG 2650

Db       2468 TCGGCGTGCCCGGTGAGATCCACGTCGGCGGCCCCGGCGTGGCACGCGGCTACCTCAACC 2527

Qy       2651 ATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGA------GACCGGGGAGCGCC 2704

Db       2528 GGCCGGAGCTGACGGCCGAGCGGTTCGTCCCGGACCCGTACGGTCCGGCGGGCTCGCGTC 2587

Qy       2705 TCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGC 2764

Db       2588 TCTACAAGAGCGGCGACCTCGCGCGCCGGCTGCCCGACGGCAGCCTGGAGTTCCTGGGCC 2647

Qy       2765 GTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAA 2824

Db       2648 GGATCGACGACCAGGTCAAGATCCGCGGCTTCCGGATCGAGCTGGGCGAGATCGAGACCG 2707

Qy       2825 CGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGAACGACGCGG 2884

Db       2708 CGCTGGCCGCGCATCCGCAGGTCCGTGACGCCGTCGTCCTGGTCCGCGAGGACACCCCCG 2767

Qy       2885 CGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCCG 2941

Db       2768 GCGACAAGCGCCTCGTCGCCTACACGACGCCGGCCGCCGACCAGGCCCCGGCCCCCG 2824
```

RESULT 13
US-09-940-316B-1
; Sequence 1, Application US/09940316B
; Publication No. US20030175901A1
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH
; TITLE OF INVENTION: GENE CLUSTER
; FILE REFERENCE: 30062-20026.11
; CURRENT APPLICATION NUMBER: US/09/940,316B
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/410,551
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275)...(71465)
US-09-940-316B-1

```
  Query Match             5.7%;  Score 241.6;  DB 12;  Length 77536;
  Best Local Similarity   49.7%;  Pred. No. 8.5e-53;
  Matches  703;  Conservative    0;  Mismatches  699;  Indels   12;  Gaps    3;

Qy       1530 GGCGAGCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGC 1589

Db      49194 GGTGCCCTCGAACGACACGGCGCGGGACCTGCCCGTCGACACGCTGCCGGGCCTGCTGGC 49253

Qy       1590 GGCGCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCAC 1649

Db      49254 CCGGTACGCCGCACGCACCCCCGGCGCCGTGGCCGTCACCGACCCGCACATCTCCCTCAC 49313

Qy       1650 GTACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGC 1709

Db      49314 CTACGCGCAGCTGGACCGGCGGGCGAACCGCCTCGCGCACCTGCTCCGCGCGCGCGGCAC 49373

Qy       1710 ACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGT 1769

Db      49374 CGCCACCGGCGACCTGGTCGGGATCTGCGCCGATCGCGGCGCCGACCTGATCGTCGGCAT 49433

Qy       1770 TCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGA 1829

Db      49434 CGTGGGGATCCTCAAGGCGGGCGCCGCTTATGTGCCGCTGGACCCCGAACATCCTCCGGA 49493

Qy       1830 GCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCT 1889

Db      49494 GCGCACGGCGTTCGTGCTGGCCGACGCGCAGCTGACCACGGTGGTGGCGCA------CGA 49547

Qy       1890 GGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGT 1949

Db      49548 GGTCTACCGTTCCCGGTTCCCCGATGTGCCGCACGTGGTGGCGTTGGACGACCCGGAGCT 49607

Qy       1950 CGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTA 2009

Db      49608 GGACCGGCAGCCGGACGACACGGCGCCGGACGTCGAGCTGGACCGGGACAGCCTCGCCTA 49667

Qy       2010 TGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGG 2069

Db      49668 CGCGATCTACACGTCCGGGTCGACCGGCAGGCCGAAGGCCGTGCTCATGCCGGGTGTCAG 49727

Qy       2070 TGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGT 2129

Db      49728 CGCCGTCAACCTGCTGCTCTGGCAGGAGCGCACGATGGGCCGCGAGCCGGCCAGCCGCAC 49787
```

```
Qy       2130 GCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGC 2189
                   |   | ||    |||  |  |||||   ||||||  | ||    | ||        |  |
Db      49788 CGTCCAGTTCGTGACGCCCACGTTCGACTACTCGGTGCAGGA---GATCTTTTCCGCGCT 49844

Qy       2190 GGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGC 2249
                |  |||||| ||| ||||  | |||     |            || |   ||| ||| ||          ||
Db      49845 GCTGGGCGGCACGCTCGTCATCCCGCCGGACGAGGTGCGGTTCGACCCGCCGGGACTCGC 49904

Qy       2250 AGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGAT 2309
                    || ||| ||      | ||   | | ||      |          || ||   | ||||   |    ||||
Db      49905 CCGGTGGATGGACGAACAGGCGATTACCCGGATCTACGCGCCGACGGCCGTACTGCGCGC 49964

Qy       2310 GCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCT 2369
                ||| ||||||| | ||      |  |  |||   ||||||          | ||  || |  |
Db      49965 GCTGATCGAGCACGTCGATCCGCACAGCGACCAGCTCGCCGCCCTGCGGCACCTGTGCCA 50024

Qy       2370 GCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGG 2429
                |     |||   |      |||||  |    |   |  |      |||||||               |  |||||
Db      50025 GGGCGGCGAGGCGCTGATCCTCGACGCGCGGTTGCGCGAGCTGTGCCGGCACCGGCCCCA 50084

Qy       2430 CGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCC 2489
                | ||    |||   ||  |          |||   |   |||||         | ||      ||                  |
Db      50085 CCTGCGCGTGCACAATCACTACGGTCCGGCCGAAAGCCAGCTCATCACCGGGTACACGCT 50144

Qy       2490 CGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCA 2549
                        ||   ||||        | |      || | |        ||      |||||    |||    |     ||||
Db      50145 GCCCGCCGACCCCGACGCGTGGCCCGCCACCGCACCGATCGGCCCGCCGATCGACAACAC 50204

Qy       2550 GACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACT 2609
                       ||||  ||||||| |||||| |        |||        |||| | |    | |||||||| ||
Db      50205 CCGCATCCATCTGCTCGACGAGGCGATGCGGCCGGTTCCGGACGGTATGCCGGGGCAGCT 50264

Qy       2610 CTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAA 2669
                || |  |  |  ||| ||||| || ||  |   || |||     |       |        |||   |||      |  |
Db      50265 CTGCGTCGCCGGCGTCGGCCTCGCCCGTGGGTACCTGGCCCGTCCCGAGCTGACCGCCGA 50324

Qy       2670 GAGCT---TCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGG 2726
                | |||       |  |   |       |    |  |       || |||||||  |  |||      |||||||| ||||
Db      50325 GCGCTGGGTGCCGGGAGATGCGGTCGGCGAGGAGCGCATGTACCTCACCGGCGACCTGGC 50384

Qy       2727 CCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCT 2786
                ||||   |   |||||| ||    || |  || |||  || |  | ||         ||| ||||    ||||| |
Db      50385 CCGCCGCGCGCCCGACGGCGACCTGGAATTCCTCGGCCGGATCGACGACCAGGTCAAGAT 50444

Qy       2787 TCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGT 2846
                 |||||    |||||| || |    || || ||||||         ||||         | |   |     |||
Db      50445 CCGCGGCATCCGCGTCGAACCGGGTGAGATCGAGAGCCTGCTCGCCGAGGACGCCCGCGT 50504

Qy       2847 ACGCGACGCGGTGATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTA 2906
                       | |||| |  |                | |  | ||        ||      |  |||  ||||         || ||
Db      50505 CACGCAGGCGGCGGTGTCCGTGCGCGAGGACCGGCGGGGCGAGAAGTTCCTGGCCGCGTA 50564

Qy       2907 TGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
                 || || |||| || |       |||        |||    | |
Db      50565 CGTCGTACCGGTGGCCGGCCGGCACGGCGACGAC 50598
```

RESULT 14
US-09-976-059-1
; Sequence 1, Application US/09976059
; Patent No. US20020164747A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
; FILE REFERENCE: 3019-PCT
; CURRENT APPLICATION NUMBER: US/09/976,059
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 88421
; TYPE: DNA
; ORGANISM: Actinoplanes sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2077)..(3078)
; OTHER INFORMATION: ORF 1; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (3118)..(4032)
; OTHER INFORMATION: ORF 2; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (4038)..(5048)
; OTHER INFORMATION: ORF 3; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (4038)..(5048)
; OTHER INFORMATION: ORF 3; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (6665)..(5814)
; OTHER INFORMATION: ORF 4; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (7703)..(6693)
; OTHER INFORMATION: ORF 5; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (9464)..(8130)
; OTHER INFORMATION: ORF 6; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (9691)..(10761)
; OTHER INFORMATION: ORF 7; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (12751)..(10829)
; OTHER INFORMATION: ORF 8; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (13617)..(12802)
; OTHER INFORMATION: ORF 9; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (15203)..(13614)
; OTHER INFORMATION: ORF 10; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (15591)..(15863)
; OTHER INFORMATION: ORF 11; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (15880)..(19035)
; OTHER INFORMATION: ORF 12; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (19032)..(39713)
; OTHER INFORMATION: ORF 13; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (39713)..(65800)
; OTHER INFORMATION: ORF 14; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (65826)..(66530)
; OTHER INFORMATION: ORF 15; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (66546)..(67370)
; OTHER INFORMATION: ORF 16; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (67384)..(70059)
; OTHER INFORMATION: ORF 17; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (70099)..(70662)
; OTHER INFORMATION: ORF 18; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (70659)..(71906)
; OTHER INFORMATION: ORF 19; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (73439)..(71964)
; OTHER INFORMATION: ORF 20; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (74216)..(73563)
; OTHER INFORMATION: ORF 21; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (75424)..(74213)
; OTHER INFORMATION: ORF 22; negative strandedness
; NAME/KEY: misc_feature

```
;   LOCATION: (75535)..(76464)
;   OTHER INFORMATION: ORF 23; positive strandedness
;   NAME/KEY: misc_feature
;   LOCATION: (78110)..(76449)
;   OTHER INFORMATION: ORF 24; negative strandedness
;   NAME/KEY: misc_feature
;   LOCATION: (79864)..(78107)
;   OTHER INFORMATION: ORF 25; negative strandedness
;   NAME/KEY: misc_feature
;   LOCATION: (81624)..(79861)
;   OTHER INFORMATION: ORF 26; negative strandedness
;   NAME/KEY: misc_feature
;   LOCATION: (81909)..(81682)
;   OTHER INFORMATION: ORF 27; negative strandedness
;   NAME/KEY: misc_feature
;   LOCATION: (82346)..(82062)
;   OTHER INFORMATION: ORF 28; negative strandedness
;   NAME/KEY: misc_feature
;   LOCATION: (82587)..(84446)
;   OTHER INFORMATION: ORF 29; positive strandedness
;   NAME/KEY: misc_feature
;   LOCATION: (84481)..(85548)
;   OTHER INFORMATION: ORF 30; positive strandedness
;   NAME/KEY: misc_feature
;   LOCATION: (85556)..(86845)
;   OTHER INFORMATION: ORF 31; positive strandedness
;   NAME/KEY: misc_feature
;   LOCATION: (87372)..(86803)
;   OTHER INFORMATION: ORF 32; positive strandedness
;   NAME/KEY: misc_feature
;   LOCATION: (87494)..(88420)
;   OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-09-976-059-1

  Query Match             5.6%;  Score 239;  DB 10;  Length 88421;
  Best Local Similarity   50.4%;  Pred. No. 4.2e-52;
  Matches  729;  Conservative    0;  Mismatches  695;  Indels   23;  Gaps    5;

Qy       1525 GCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTG 1584
Db      68266 GCACCCGCACCCGCACCTGCGCCCGCGCCGCCGGCTGCGGCTCGCACGCTGACCGGGCTC 68325

Qy       1585 TTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACG 1644
Db      68326 TTCGCCGAGCAGGTCGCCGCTCGTCCCACGGCCGTCGCCGTCTCGGACGATCGGGGTCGG 68385

Qy       1645 CTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAG 1704
Db      68386 CACACCTACCGCGAGCTCGACGAGTGGTCCGGCCGGCTGGCCCGGGGGCTGCGGAAGGCC 68445

Qy       1705 GGGGCACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTC 1764
Db      68446 GGCGTGCGCGACGGCGACGCGGTCGGCGTCTGCCTCGACCGCTCGGCCGAGCTCGTCGCC 68505

Qy       1765 GCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCG 1824
Db      68506 GTGCTCCTCGCGGTCCTCAAGGCCGGCGCCGCCTATGTGCCGCTCGACGCGGCGTACCCG 68565

Qy       1825 GCGGAGCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCA 1884
Db      68566 GCCGACCGCATCGCCTACACCGTCGGCGACGCCGGCCTCGCGGTCGTGGTCACCACCTCG 68625

Qy       1885 TGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCC 1944
Db      68626 GCGGACTTTCCCGACGTGGACGGTGTGCGGCTGCTCGCGCCGGAGAGCCTCGCCGAGGCC 68685

Qy       1945 GGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTC 2004
Db      68686 GGCGACGACCCGGGCATC---------CCGCTCGCCACCCCGGCCGGCCCGGAGCGGCCG 68736

Qy       2005 GCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCAT 2064
Db      68737 GCCTATGTCATCTACACGTCCGGTTCCACGGGCCGGCCCAAGGGCGTGGTCGTCCCGCAC 68796

Qy       2065 CGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGAC 2124
Db      68797 GCCAACGTGTCCGCGCTGCTCGACGCCACGCGCGAGGAGTACGCGCTCGGCCCCGGCGAC 68856

Qy       2125 AGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATC 2184
Db      68857 GTGTGGACCTTCTTCCACTCGGCCGCCTTCGACTTCTCCGTCTGGGAGATCTGGGGCTGC 68916

Qy       2185 CTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCAT 2244
Db      68917 CTGCTCACCGGCGGCCACCTCGTCGTCGTCCCGTACTGGGTGTCCCGCTCGCCGGAGCAG 68976

Qy       2245 TGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATG 2304
Db      68977 TTCCACGACCTGCTCGCCGAGCGCGGCGTCACCGTGCTCAACCAGACGCCGTCCAGCTTC 69036

Qy       2305 CGGATGCTCGTCG-AGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCT 2363
Db      69037 ACGCAGCTCGTGGCCGCCGACCGCGGGGCGGAGCGCGACCTCGCCGTACGCCTGGTGATT 69096

Qy       2364 TTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAG 2423
Db      69097 TTCGGTGGTGAGCCGCTCGACGCCCGGACGGTGCTGCCCTGGCTGGACCGCCGTCCCGAG 69156

Qy       2424 GCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGG 2483
Db      69157 GCGC-GCTGCCGGCTGGTCAACATGTTCGGCATCACCGAGACCACCGTGCACGTCACGGC 69215

Qy       2484 GTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCG 2543
Db      69216 GGTCGACGTCACGCGCGCGGCCGCGCTCGCCGGCTCCCGGTCGGTCGGCCGCCCGCTGCC 69275

Qy       2544 CAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGG 2603
Db      69276 CGGCTGGGCCGTGCGCGTGCTCGACGAGCAGCGCCGCGAGGTGCCGCCGGGCGTGCCGGG 69335

Qy       2604 GCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGAC 2663
Db      69336 CGAGATCTACGTGGGCGGCGCCGGCGTGGCGATCGGCTACCTCAACCGCCCGGAGCTGAC 69395

Qy       2664 GCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCT 2723
Db      69396 CGCCGAGCGGTTCGTCACCGGCCCGGA---CGGCCGGCGCTGGTACCGCTCCGGCGACCG 69452

Qy       2724 GGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAA 2783
Db      69453 CGGCCGGCTGCTGCCCGACGGCACCCTGGAACACCTGGGCCGGCTCGACGACCAGGTCAA 69512

Qy       2784 GCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAA 2843
Db      69513 GCTGCGCGGCTTCCGGATCGAGCTGGACGAGATCCGGGGCGTGCTCACCGAGTGCGCCGG 69572

Qy       2844 CGTACGCGACGCGGTGATTGT---------GCCCGTCGGGAACGACGCGGCGAACAAGCT 2894
Db      69573 GGTGGCGGCGGCCGCGGTCGTCATCCGGCGCTCCACTCCGGACGATCCGGCGACCGCGCG 69632

Qy       2895 CCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCCGAGCAGGACGCGAG 2954
Db      69633 GCTCGACGCGTACGTGGTCGCCGAGGCCGGCGCCACGCCGCCGGTGGCCGAGCACGCGGC 69692

Qy       2955 CCTCAAG 2961
Db      69693 CCGGATG 69699

RESULT 15
US-10-156-761-3630
; Sequence 3630, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
```

```
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3630
;   LENGTH: 5127
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(5127)
US-10-156-761-3630

  Query Match             5.5%;  Score 233.4;  DB 14;  Length 5127;
  Best Local Similarity   51.0%;  Pred. No. 7.1e-51;
  Matches  692;  Conservative    0;  Mismatches  641;  Indels   24;  Gaps    5;

Qy       1559 TGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGC 1618
Db          2 TGAACGCCGAACTGCTTTCGGAACGCTTCGCCGAGCAGGCGTTCCGTACGCCGCACGCGG 61

Qy       1619 TCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGC 1678
Db         62 CCGCCGTCGTCGACGGAGCCCGGACGGTGACGTACGAGGAGCTGGACCGTGCGTCCCTGC 121

Qy       1679 GACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGA 1738
Db        122 GCATGGCCCGCCACCTGCGGGACCTGGGTGCGGGGCCCGAGACGCTCGTCGGGGTGAGCC 181

Qy       1739 TGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCT 1798
Db        182 TGCCGCGCGGCGTCGATCTGATCGTCGCGCTGCTGGCCGTCTGGCGGGCCGGCGCGGGCT 241

Qy       1799 ACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTG 1858
Db        242 ATGTCCCGCTCGACCCGGCCCAGCCGCCCGCCCGCCTGTCGGACCTGGCCCGGGAGGCCG 301

Qy       1859 AGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGA 1918
Db        302 GGGCGCGCCTGGTGGTCGCCGGACCGGCCCTGGCGGGC--CCCGTGCGCGACGCCGGTGC 359

Qy       1919 TCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGA 1978
Db        360 ----CCGGCGCGTCGGACCCGAGGAGATCCCGGACCGGTCCTCGGACCCCGCCGGCCCGC 415

Qy       1979 TGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGT 2038
Db        416 TCCCCGCCGCCGACCCCGCCAACGCGGCGTACGCCGTCTTCACCTCCGGTTCGACCGGCC 475

Qy       2039 TGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACG 2098
Db        476 GGCCCAAGGCCGTCGTGGTCACCCACGCGGGCATCGCGAACCGGATCGGCTGGACCGTGC 535

Qy       2099 AGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATC 2158
Db        536 AGCGGCATGCGCTCGGGGCGGCCGACCGGGTTCTGCAGAAGACCACCATCGGGTTCGACG 595

Qy       2159 TCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACG 2218
Db        596 CGGCCGGCTGGGAGATCTTCGCGCCGCTGGTCGGCGGCGGCACCGTGGTGCTCGCCCCGG 655

Qy       2219 CGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGG 2278
Db        656 CGGGCGCCGAGCGCGACCCGGCCGCGCTGCTGCGCGCGGTGGCCGACCACGGCGTCACCG 715

Qy       2279 TGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTGAGGGTCGCCCCG 2338
Db        716 TGCTCCAGGTCGTGCCGTCGGTGCTGCGGCTGCTCGTGGAGGAGGGTGACTGGAGCGGC- 774

Qy       2339 ATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCC 2398
Db        775 -----TGCGGCTCGCTGCGGCTGCTGTTCTCGGCGGGCGAGGCCCTGCACGCCGAGCTGG 829

Qy       2399 TGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCA 2458
Db        830 TCGCGCGGCTGCGTGAGCGGACGGGCCAGGACCTGGAGGTCTGGAACACCTACGGTCCCA 889

Qy       2459 CCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGA 2518
Db        890 CCGAGTGCTCCATCGACATCACCGCGCAGCTCGT---CGACCCCGCGCTCACCGCGGGTC 946

Qy       2519 GCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCG 2578
Db        947 CGGTGCCCATCGGCCGCCCGCTGCCCGGCATGCGCGTCCTCGTCCTCGGACCGAACGGTG 1006

Qy       2579 AACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGG 2638
Db       1007 TCCCCGTGCCCGTCGGTGTTCCCGGCGAGCTGTACGCGGGCGGCGTCGGAGTGGCCCGCG 1066

Qy       2639 GCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACC---------CCG 2689
Db       1067 GCTATGCCGGGCGCCCCGATCTGACCGCCGACCGTTTCGTGCCCGACCCCTACGCCACCG 1126

Qy       2690 AGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACA 2749
Db       1127 AGCCCGGCGCCCGCCTCTACCGCACCGGTGACCAGGTGCGCTGGCGCTCCGACCGCACCC 1186

Qy       2750 TCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCG 2809
Db       1187 TCGAGTACCTGGGCCGGCTCGACCACCAGGTGAAGGTCAACGGCGTCCGGATCGAGCCCG 1246

Qy       2810 GGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCG 2869
Db       1247 CCGAGGTCGAGGCCGCGCTCGCCGCGCACCCGGCGGTCACCGGCGCGGTCGTCACGCCGT 1306

Qy       2870 TCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTA 2906
Db       1307 ACGAGGCCGACGGCGGCGGCAAGCGGCTCGCCGCGTA 1343

Search completed: October  4, 2003, 12:40:37
Job time : 1029.11 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:         October 3, 2003, 11:46:26 ; Search time 224.251 Seconds
                                           (without alignments)
                                           8331.640 Million cell updates/sec

Title:          US-10-014-717-1_COPY_11872_16104
Perfect score:  4233
Sequence:       1 atgacgatcaatcagcttct..........agggcaggagacgtagctaa 4233

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters:     1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4233	100.0	68750	3	US-09-335-409-1	Sequence 1, Appli
	2	4233	100.0	68750	4	US-09-568-102-1	Sequence 1, Appli
	3	4233	100.0	68750	4	US-09-567-969-1	Sequence 1, Appli
	4	4233	100.0	68750	4	US-09-568-480-1	Sequence 1, Appli
	5	4233	100.0	68750	4	US-09-568-486-1	Sequence 1, Appli
	6	4233	100.0	68750	4	US-09-568-472-1	Sequence 1, Appli
	7	4233	100.0	68750	4	US-09-567-899-1	Sequence 1, Appli
	8	4114.6	97.2	71989	4	US-09-443-501A-2	Sequence 2, Appli
	9	520.6	12.3	4563	4	US-09-252-991A-4765	Sequence 4765, Ap
c	10	491	11.6	2865	4	US-09-252-991A-4675	Sequence 4675, Ap
	11	455	10.7	2922	4	US-09-252-991A-4815	Sequence 4815, Ap
c	12	358.2	8.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	13	349.2	8.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	14	348.6	8.2	1842	4	US-09-252-991A-4678	Sequence 4678, Ap
	15	340	8.0	2844	4	US-09-252-991A-4764	Sequence 4764, Ap
c	16	274.4	6.5	3315	4	US-09-252-991A-8892	Sequence 8892, Ap
	17	274.4	6.5	7911	4	US-09-252-991A-9182	Sequence 9182, Ap
c	18	261.4	6.2	4236	4	US-09-252-991A-7057	Sequence 7057, Ap
	19	261.4	6.2	10023	4	US-09-252-991A-6997	Sequence 6997, Ap
	20	246.6	5.8	6858	4	US-09-252-991A-1219	Sequence 1219, Ap
	21	241.6	5.7	77536	4	US-09-410-551B-1	Sequence 1, Appli
	22	240.8	5.7	1320	4	US-09-252-991A-4790	Sequence 4790, Ap
	23	229.8	5.4	6573	4	US-09-252-991A-9183	Sequence 9183, Ap
c	24	229.6	5.4	1857	4	US-09-252-991A-8888	Sequence 8888, Ap
	25	226.6	5.4	630	4	US-09-252-991A-4792	Sequence 4792, Ap
c	26	223.2	5.3	2025	4	US-09-252-991A-8891	Sequence 8891, Ap
	27	220.2	5.2	7374	4	US-09-252-991A-9100	Sequence 9100, Ap
c	28	218.8	5.2	1689	4	US-09-252-991A-8717	Sequence 8717, Ap
c	29	202	4.8	4242	4	US-09-252-991A-7056	Sequence 7056, Ap
	30	200.6	4.7	1965	4	US-09-252-991A-9230	Sequence 9230, Ap
	31	194	4.6	2274	4	US-09-252-991A-6998	Sequence 6998, Ap
c	32	194	4.6	2823	4	US-09-252-991A-7100	Sequence 7100, Ap
	33	194	4.6	3456	4	US-09-252-991A-7025	Sequence 7025, Ap
c	34	186	4.4	1509	4	US-09-252-991A-63	Sequence 63, Appl
	35	186	4.4	7527	4	US-09-252-991A-71	Sequence 71, Appl
	36	182.8	4.3	804	4	US-09-252-991A-4819	Sequence 4819, Ap
	37	175.4	4.1	954	4	US-09-252-991A-4820	Sequence 4820, Ap
	38	169.2	4.0	2805	4	US-09-252-991A-9064	Sequence 9064, Ap
	39	168.2	4.0	2219	3	US-08-510-646B-17	Sequence 17, Appl
	40	166.2	3.9	1494	4	US-09-252-991A-7049	Sequence 7049, Ap
c	41	163.2	3.9	765	4	US-09-252-991A-4721	Sequence 4721, Ap
c	42	158.2	3.7	1178	3	US-08-861-774E-91	Sequence 91, Appl
	43	155	3.7	1200	3	US-08-861-774E-85	Sequence 85, Appl
	44	152.8	3.6	1299	4	US-09-252-991A-7051	Sequence 7051, Ap
	45	151.4	3.6	3126	4	US-09-252-991A-9066	Sequence 9066, Ap

ALIGNMENTS

```
RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-335-409-1

  Query Match             100.0%;  Score 4233;  DB 3;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231
```

```
Qy      361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy      421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy      481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy      541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy      601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy      661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy      721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy      781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy      841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy      901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy      961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy     1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy     1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy     1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy     1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy     1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy     1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy     1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy     1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy     1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy     1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy     1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy     1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy     1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy     1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy     1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy     1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy     1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy     2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy     2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy     2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy     2221 TCCAAGCTGCGCGATCCGGCGCATTGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14092 TCCAAGCTGCGCGATCCGGCGCATTGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy     2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy     2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy     2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy     2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy     2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy         2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy         2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy         2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy         2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy         2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy         2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy         2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy         3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy         3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy         3121 GACGTCTACGCGCGTCGCCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14992 GACGTCTACGCGCGTCGCCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy         3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy         3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy         3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy         3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy         3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy         3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy         3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy         3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy         3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy         3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy         3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy         3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy         3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy         3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy         4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy         4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy         4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy         4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
                |||||||||||||||||||||||||||||||||
Db        16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

```
  Query Match             100.0%;  Score 4233;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy            1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
```

```
Db     11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy        61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
Db     11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy       121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
Db     11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy       181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
Db     12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy       241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
Db     12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy       301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
Db     12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy       361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
Db     12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy       421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
Db     12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy       481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
Db     12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy       541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
Db     12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy       601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
Db     12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy       661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
Db     12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy       721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
Db     12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy       781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
Db     12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy       841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
Db     12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy       901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
Db     12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy       961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
Db     12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy      1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
Db     12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy      1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
Db     12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy      1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
Db     13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy      1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
Db     13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy      1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
Db     13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy      1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
Db     13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy      1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
Db     13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy      1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
Db     13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy      1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
Db     13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy      1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
Db     13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy      1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
Db     13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy      1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
Db     13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy      1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
Db     13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy      1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
Db     13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy      1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
Db     13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy      1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
Db     13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy      1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
Db     13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy      2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
Db     13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy      2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
Db     13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy      2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
Db     14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091
```

```
Qy       2221 TCCAAGCTGCGCGATCCGGCGCATTGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14092 TCCAAGCTGCGCGATCCGGCGCATTGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy       2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy       2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy       2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy       2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy       2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy       2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGCTGGCACTGGGC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGCTGGCACTGGGC 14511

Qy       2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy       2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy       2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy       2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy       2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy       2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy       3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy       3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy       3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy       3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy       3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy       3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy       3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||||||||||||||||||||||||||||||||
Db      16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan

```
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             100.0%;  Score 4233;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy       1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy       1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy       1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy       1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy       1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy       1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy       1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy       1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy       1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy       1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy       1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy       1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy       1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy       2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy       2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy       2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy       2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy       2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy       2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy       2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy       2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy       2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy       2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy       2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy       2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy       2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy       2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy       2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy       2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy       3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy       3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy       3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy       3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy       3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy       3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy       3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951
```

```
Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||||||||||||||||||||||||||||||||
Db      16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

```
RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             100.0%;  Score 4233;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy       1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy       1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy       1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy       1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy       1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431
```

```
Qy       1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy       1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy       1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy       1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy       1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy       1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy       1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy       1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy       2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy       2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy       2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy       2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy       2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy       2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy       2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy       2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy       2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy       2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy       2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy       2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy       2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy       2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy       2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy       2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy       3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy       3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy       3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy       3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy       3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy       3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy       3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||||||||||||||||||||||||||||||||
Db      16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104


RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             100.0%;  Score 4233;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy         1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy         1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy         1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy         1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy         1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy         1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy         1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy         1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy         1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy         1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy         1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy         1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy         1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy         2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy         2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy         2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy         2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy         2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy         2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy         2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy         2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy         2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy         2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy         2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy         2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy         2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy         2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy         2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy         2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy         3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy         3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy         3121 GACGTCTACGCGCGTCGCCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14992 GACGTCTACGCGCGTCGCCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy         3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy         3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy         3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy         3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291
```

```
Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||||||||||||||||||||||||||||||||
Db      16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

```
  Query Match             100.0%;  Score 4233;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771
```

```
Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy       1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy       1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy       1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy       1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy       1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy       1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy       1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy       1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy       1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy       1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy       1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy       1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy       1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy       2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy       2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy       2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy       2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy       2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy       2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy       2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy       2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy       2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy       2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy       2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy       2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy       2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy       2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy       2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy       2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy       3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy       3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy       3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy       3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy       3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy       3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy       3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||||||||||||||||||||||||||||||||
Db      16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

```
  Query Match             100.0%;  Score 4233;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
```

```
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy          601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy          661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy          721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy          781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy          841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy          901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy          961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy         1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy         1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy         1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy         1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy         1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy         1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy         1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy         1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy         1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy         1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy         1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy         1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy         1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy         1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy         1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy         1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy         1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy         2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy         2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy         2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy         2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy         2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy         2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy         2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy         2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy         2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy         2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy         2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy         2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631
```

```
Qy       2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy       2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy       2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy       2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy       3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy       3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy       3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy       3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy       3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy       3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy       3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||||||||||||||||||||||||||||||||
Db      16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

```
  Query Match             97.2%;  Score 4114.6;  DB 4;  Length 71989;
  Best Local Similarity   98.3%;  Pred. No. 0;
  Matches 4159;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       6260 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTGTCAAGCTGGCGGCCGAT 6319

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       6320 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 6379

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
Db       6380 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAGTCC 6439

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
Db       6440 ATCGTGCCCGCCCCAGCCGAGCGGCACGTTCCGTTTCCTCTCACAGACATCCAAGGATCC 6499

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
Db       6500 TACTGGCTGGGTCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 6559

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
Db       6560 TACGACTGTACGGATCTCGACGTGGCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 6619

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
Db       6620 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 6679

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
Db       6680 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 6739

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
Db       6740 AGGCTCGTATCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 6799

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
Db       6800 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCAGCAAACCCGTCTCGTGCTCAGTATC 6859

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
Db       6860 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGATTGGCTCAGCTTC 6919

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
Db       6920 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGACTATGTGCTC 6979

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
Db       6980 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 7039

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
Db       7040 CGCGTCGCCGAGCTCCCACCTCCGCCGATGCTTCCGATGAAGGCCGATCCATCTACCCTG 7099

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
Db       7100 AGGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGAGTCGATTG 7159

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
Db       7160 AAGCAGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATTCTGGCTGCATTTTCC 7219

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
Db       7220 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 7279

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
Db       7280 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 7339

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
Db       7340 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 7399

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
Db       7400 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 7459

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
Db       7460 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTC 7519

Qy       1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
Db       7520 AACCAGCAAGTCGTTGGTGTCACCTCGCTGCAGAGGCTCGGCACTCCGGTGTACACCAGC 7579

Qy       1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
Db       7580 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 7639

Qy       1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
Db       7640 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 7699

Qy       1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
Db       7700 GCGTACGTCGCTTTTCTCCGGCGGCTCACTGAGGAACCATGGAGTGAACAGATGCGCTGT 7759

Qy       1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
Db       7760 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGAGACCAACTCGCTGCTG 7819

Qy       1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
Db       7820 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCTATGCAGCTC 7879

Qy       1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
Db       7880 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 7939

Qy       1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
Db       7940 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 7999

Qy       1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
Db       8000 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 8059

Qy       1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
Db       8060 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 8119

Qy       1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
Db       8120 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 8179

Qy       1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
Db       8180 CAGCGGCTGCTCGTGAGCGATGCCGGCGTCGAAGGCGACGGCGACCAGCTTCCGATGATG 8239

Qy       1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
Db       8240 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 8299

Qy       2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
Db       8300 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 8359

Qy       2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
Db       8360 CGCTTCGAAATAGGGCCCGGAGACAGAGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 8419

Qy       2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
Db       8420 TCGGTCTACGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 8479

Qy       2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
Db       8480 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGCGTTGATCGAACGAGAGAAGGTGACGGTG 8539
```

```
Qy        2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
Db        8540 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTCCGAGGGTCGCCCCGAT 8599

Qy        2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
Db        8600 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 8659

Qy        2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
Db        8660 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 8719

Qy        2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
Db        8720 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGATCCATCGTGGGCGAGC 8779

Qy        2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
Db        8780 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 8839

Qy        2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGCTGGCACTGGGC 2640
Db        8840 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGACTGGCACTGGGC 8899

Qy        2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
Db        8900 TACTGGCGCGATGAAGAGAAGACGCGCAACAGCTTCCTCGTGCACCCCGAGACCGGGGAG 8959

Qy        2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
Db        8960 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 9019

Qy        2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
Db        9020 GGGCGGGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 9079

Qy        2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
Db        9080 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 9139

Qy        2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
Db        9140 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAAGGCACACGGAGACGCGCTGCC 9199

Qy        2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
Db        9200 GAGCAGGACGCGAGCCTCAAGACCGAGCGGGTCGACGCGAGAGCACACGCCGCCAAAGCG 9259

Qy        3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
Db        9260 GACGGATTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 9319

Qy        3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
Db        9320 GATCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCTGGTTCCGCGGGAGGCGGGGCTG 9379

Qy        3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
Db        9380 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTCGAGGCCCCGATTCCATTTGTT 9439

Qy        3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
Db        9440 GAATTCGGCCGATTCCTGAGCTGCCTGAGCAGCGTGGAGCCCGACGGCGCGGCCCTTCCC 9499

Qy        3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300.
Db        9500 AAATTCCGTTATCCATCGGCTGGCAGCACGTACCCGGTGCAAACCTACGCGTACGCCAAA 9559

Qy        3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
Db        9560 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 9619

Qy        3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
Db        9620 TTGCTGAAGGTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCCGCAAAACTTCGAC 9679

Qy        3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
Db        9680 GTGTTCGATGAAGCGGCGTTCGGCCTCCTGTTCGTGGGCAGGATCGATGCCATCGAGTCG 9739

Qy        3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
Db        9740 CTGTATGGATCGTTGTCACGAGAATTCTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 9799

Qy        3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
Db        9800 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGTCAATTCGATTTT 9859

Qy        3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
Db        9860 GAACAGGTTCGGCCGGTTCTCGACCTGCGGCATTCGGACGTTTACGTGCACGGCATGCTG 9919

Qy        3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
Db        9920 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 9979

Qy        3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
Db        9980 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGATCAGCACTTCGCCGATATCCTT 10039

Qy        3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
Db       10040 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 10099

Qy        3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
Db       10100 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 10159

Qy        3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
Db       10160 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 10219

Qy        3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
Db       10220 CTCGTTGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTTGGGCTCCAGCAGAGCTTC 10279

Qy        4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
Db       10280 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGTCTGTTGCAGAAGAGG 10339

Qy        4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
Db       10340 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 10399

Qy        4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
Db       10400 TCCGGTTTGCGCCGAGACTCGAAAGATCTAGAGCAGCGGCCGAACATGCAGGACCGAGTG 10459

Qy        4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
Db       10460 GAGGCTCGGCGCAAGGGCAGGAGACGTAGCTAA 10492
```

RESULT 9
US-09-252-991A-4765
; Sequence 4765, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

```
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4765
;   LENGTH: 4563
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4765

  Query Match             12.3%;  Score 520.6;  DB 4;  Length 4563;
  Best Local Similarity   52.0%;  Pred. No. 3.5e-105;
  Matches 1511;  Conservative    0;  Mismatches 1329;  Indels   63;  Gaps   13;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
Db        124 ATGAGCCTCGGCGAACTGCTGGAAACCTGCCGCAGCCGGCGCATCGAACTCTGGAGCGAG 183

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
Db        184 GCGGGCCGCCTGCGCTATCGCGCCCCCCAGGGCGCCCTCGACGCCGGCCTCGCCGAGCG- 242

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
Db        243 --CCTGCGGGCCGAGCGCGAGGCCCTGCTGGAACACCTGGAAGGCGGCCCTGGCTGGCGC 300

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
Db        301 GCCGAACCCGACCTGGCCCACCAGCGC------TTCCCATTGACCCCGGTGCAGGCCGCC 354

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTT---TACGGTCCCCAGCGGGATCCACGCCTATCGC 297
Db        355 TACGTGCTGGGCCGCCAGGCGGCCTTCGACTACGGCGGTAACGCCTGCCAGCTGTACGCC 414

Qy        298 GAATACGACTGTACG---GATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTC 354
Db        415 GAGTACGACTGGCCGGTCGACACCGATCCGGCGCGCCTGGAGGCGGCCTGGAACGCCATG 474

Qy        355 GTCGCGCGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAG 414
Db        475 GTCGAGCGCCACCCGATGCTGCGCGCGGTGATCGAGGACAACGCCTGGCAGCGCGTGCTG 534

Qy        415 CCTAAAGTCGACGCCGACATCGAGATCATCGATCTGCGC---GGGCTCGACCGGAGCACA 471
Db        535 CCCGAGGTGCCCTGGCAGCGGCTGACCGTGCATGCCTGCGCGGGGCTCGACGAGGCCGCT 594

Qy        472 CGGGAAGCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAG 531
Db        595 TTCCAGGCGCACCTGGAGCGGGTCCGCGAACGCCTCGACCACGCCTGCGCGGCGCTCGAC 654

Qy        532 CGCCCTCCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTG 591
Db        655 CAGTGGCCGGTC---CTGCGCCCCGAGCTGAGTATCGGCCGGGATGCCTGCGTACTGCAC 711

Qy        592 CTCAGTATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGG 651
Db        712 TGCTCGGTGGATTTCACCCTGGTCGACTACGCCAGCCTGCAATTGCTGCTTGGCGAATGG 771

Qy        652 CTCAGCTTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGAT 711
Db        772 CGCCGCCGCTATCTCGATCCGCAATGGACGGCGGAACCGCTGGAGGCGACCTTCCGCGAC 831

Qy        712 TATGTACTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTAC 771
Db        832 TATGTCGGCGTCGAGCAGCGCCGACGCCAGTCGCCAGCCTGGCAGCGCGACCGCGACTGG 891

Qy        772 TGGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCA 831
Db        892 TGGCTGGCGCGTCTCGACGCGCTACCGGGGCGTCCCGACCTGCCGCTGCGGGTGCAGCCG 951

Qy        832 TCTACCCTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGG 891
Db        952 GACACCC---GGTCCACGCGCTTCCGGCACTTCCACGCGCGCCTCGACGAGGCCGCCTGG 1008

Qy        892 GGTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCT 951
Db       1009 CAGGCGCTCGGCGCGCGCGCCGGCGAACACGGCCTGAGCGCTGCCGGCGTGGCCTTGGCG 1068

Qy        952 GCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACG 1011
Db       1069 GCCTTCGCCGAGACCATCGGTCGCTGGAGCCAGGCACCGGCGTTCTGTCTCAACCTGACG 1128

Qy       1012 CTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCG 1071
Db       1129 GTACTCAACCGGCCGCCGCTGCATCCGCAGCTGGCGCAGGTGCTCGGTGACTTCACCGCG 1188

Qy       1072 ATGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGT 1131
Db       1189 CTCAGCCTGCTGGCAGTGGACAGCCGCCACGGCGACAGTTTCGTCGAGCGTGCCCGACGC 1248

Qy       1132 ATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAG 1191
Db       1249 ATCGGCGAGCAGATGTTCGACGACCTCGACCACCCGACCTTCAGCGGCGTCGACCTGCTG 1308

Qy       1192 CGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACG 1251
Db       1309 CGCGAACTGGCGCG---CCGGCGCGGTCGCGGCGCCGACCTGATGCCGGTGGTGTTCACC 1365

Qy       1252 AGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTG 1311
Db       1366 AGTGGCATCGGCAGCGTGCAGCGCCTGCTCGGCGATGGCGAGGCGCCGCGCGCGCCACGC 1425

Qy       1312 TACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGG 1371
Db       1426 TACATGATCAGCCAGACCCCGCAGGTCTGGCTGGACTGCCAGGTCACCGACCAGTTCGGC 1485

Qy       1372 GACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGAC 1431
Db       1486 GGCCTGGAGATCGGCTGGGACGTACGCCTCGGGTTGTTCCCCGAGGGCCAGGCGGAAGCC 1545

Qy       1432 ATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAG 1491
Db       1546 ATGTTCGACGACTTCGTCGGGCTGCTCCGGCGCCTGGCGCAGAGCCCGCGCGCCTGGACC 1605

Qy       1492 GTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAAC 1551
Db       1606 GACGGCGATGCCACGGAACCCGTCGAGGCGCCGCCGCAGGCG------------------ 1647

Qy       1552 GCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCC 1611
Db       1648 TTGCCCGGTAGTGCCCGGAGCATCGCCGCCGGTTTCGCCGAGCGTGCCCTGCTGACCCCC 1707

Qy       1612 ATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGT 1671
Db       1708 GACGCCACGGCGATCCACGATGCCGCCGGCAGCTACAGCTACCGCCAGGTCGCCCAGCAC 1767

Qy       1672 TCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCG 1731
Db       1768 GCCACCGCCCTGCGCCGCGTCCTGGAAGCGCACGGCGCGGGCCGTGGCCGGCGGGTCGCG 1827

Qy       1732 GTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGC 1791
Db       1828 GTGATGCTGCCGAAAAGCGCCGCGCAATTGGTCGCGGTGATCGGCATCCTCCAGGCCGGC 1887

Qy       1792 GCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGAT 1851
Db       1888 GCCGCCTATGTGCCGGTGGACATCCGCCAGCCTCCGCTGCGGCGCCAGGCGATCCTCGCC 1947

Qy       1852 CATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCG 1911
Db       1948 AGCGCCGAA---------GTGGTCGCGCTGGTCTGCCTGGAAAGCGA--TGTCCCGGACG 1996

Qy       1912 CCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCT 1971
Db       1997 TCGGCTGCGCCTGCGTGGCCATCGACCGGCTG----GCCGCCGACAGCGCCTGGCCGCCA 2052

Qy       1972 CCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCC 2031
```

```
Db       2053 CCGCCCGCGGCGGAGGTGGCGGCGGACGACCTCGCCTACGTGATCTACACCTCCGGCTCC 2112

Qy       2032 ACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGAC 2091
Db       2113 ACCGGCACGCCAAAGGGCGTGATGCTCAGCCATGCGGCGGTGAGCAACACGCTGCTCGAC 2172

Qy       2092 ATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGC 2151
Db       2173 ATCAACCAGCGCTACGGCGTCGACGCCAACGACCGCGTCCTCGGCCTCGCCGAGCTGAGC 2232

Qy       2152 TTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGCGGTACGATCGTGGTG 2211
Db       2233 TTCGACCTCTCGGTCTACGACTTCTTCGGCGCCACCGCGGCGGGGGCCCAGGTGGTCCTC 2292

Qy       2212 CCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAG 2271
Db       2293 CCGGACCCGGCGCGCGGCAGCGATCCATCGCACTGGGCGGAACTGCTGGAACGCCACGCC 2352

Qy       2272 GTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGT 2331
Db       2353 ATCACCCTGTGGAACTCGGTGCCGGCCCAAGGCCAGATGCTCATCGATTACCTGGAGAGC 2412

Qy       2332 CGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCG 2391
Db       2413 GAGCCGCAACGTCACCTGCCGGGACCGCGCTGCGTGCTCTGGTCCGGTGACTGGATTCCG 2472

Qy       2392 GTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGC 2451
Db       2473 GTCAGCCTGCCGACCCGATGGTGGCGGCGCTGGCCGGACAGCGCGCTGTTCAGCCTGGGC 2532

Qy       2452 GGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCG 2511
Db       2533 GGCGCCACCGAGGCGGCGATCTGGTCGATCGAGCAGCCGATCCGCCCGCAGCACACCGAG 2592

Qy       2512 TGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAG 2571
Db       2593 CTGGCCAGCATCCCTTATGGCCGTGCCCTGCGCGGGCAGAGCGTGGAAGTCCTGGATGCC 2652

Qy       2572 GCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTG 2631
Db       2653 CGCGGGCGGCGCTGCCCGCCGGGCGTGCGCGGCGAGATCCATATCGGCGGGGTGGGCCTG 2712

Qy       2632 GCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAG 2691
Db       2713 GCGCTCGGCTACGCCGGCGATCCGCAGCGCACCGCCGAACGCTTCGTCCGTCACCCCGA- 2771

Qy       2692 ACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATC 2751
Db       2772 --TGGCCGTCGCCTGTATCGCACCGGCGACCTCGGCCGCTATCTGGCCGACGGCAGCATC 2829

Qy       2752 GAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGG 2811
Db       2830 GAGTTCCTCGGCCGCGAGGACGACCAGGTGAAGATTCGCGGCCACCGCATCGAACTGGCC 2889

Qy       2812 GAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTC 2871
Db       2890 GAGCTGGACGCCGCGCTGTGCGCTCATCCGCAGGTCAATCTGGCGGCCACCGTGGTGCTC 2949

Qy       2872 GGGAACGACGCGGCGAACAAGCT 2894
Db       2950 GGCGAGACCCACGAGCGCAGCCT 2972
```

RESULT 10
US-09-252-991A-4675/c
; Sequence 4675, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4675
; LENGTH: 2865
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4675

```
  Query Match             11.6%;  Score 491;  DB 4;  Length 2865;
  Best Local Similarity   54.8%;  Pred. No. 9.9e-99;
  Matches 1121;  Conservative    0;  Mismatches  885;  Indels   39;  Gaps    6;

Qy        850 CGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTGAAGCGGCGT 909
Db       2811 CGCTTCCGGCACTTCCACGCGCGCCTCGACGAGGCCGCCTGGCAGGCGCTCGGCGCGCGC 2752

Qy        910 GTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCCGAGGTGATC 969
Db       2751 GCCGGCGAACACGGCCTGAGCGCTGCCGGCGTGGCCTTGGCGGCCTTCGCCGAGACCATC 2692

Qy        970 GGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAACCGGCTCCCC 1029
Db       2691 GGTCGCTGGAGCCAGGCACCGGCGTTCTGTCTCAACCTGACGGTACTCAACCGGCCGCCG 2632

Qy       1030 GTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTCCTGGACATC 1089
Db       2631 CTGCATCCGCAGCTGGCGCAGGTGCTCGGTGACTTCACCGCGCTCAGCCTGCTGGCAGTG 2572

Qy       1090 GACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAGCAGCTGTGG 1149
Db       2571 GACAGCCGCCACGGCGACAGTTTCGTCGAGCGTGCCCGACGCATCGGCGAGCAGATGTTC 2512

Qy       1150 GAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCCGCCCGGGTC 1209
Db       2511 GACGACCTCGACCACCCGACCTTCAGCGGCGTCGACCTGCTGCGCGAACTGGCGCG---C 2455

Qy       1210 CTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTTAACCAGCAA 1269
Db       2454 CGGCGCGGTCGCGGCGCCGACCTGATGCCGGTGGTGTTCACCAGTGGCATCGGCAGCGTG 2395

Qy       1270 GTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGCACGCAGACT 1329
Db       2394 CAGCGCCTGCTCGGCGATGGCGAGGCGCCGCGCGCGCCACGCTACATGATCAGCCAGACC 2335

Qy       1330 CCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTCCTCGCGTGG 1389
Db       2334 CCGCAGGTCTGGCTGGACTGCCAGGTCACCGACCAGTTCGGCGGCCTGGAGATCGGCTGG 2275

Qy       1390 GACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAAGCGTACGTC 1449
Db       2274 GACGTACGCCTCGGGTTGTTCCCCGAGGGCCAGGCGGAAGCCATGTTCGACGACTTCGTC 2215

Qy       1450 GTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGTTCGCTTCCG 1509
Db       2214 GGGCTGCTCCGGCGCCTGGCGCAGAGCCCGCGCGCCTGGACCGACGGCGATGCCACGGAA 2155

Qy       1510 CCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTGAGCGAGCAT 1569
Db       2154 CCCGTCGAGGCGCCGCCGCAGGCGTT-----------------GCCCGGTAGTGCCCGG 2113

Qy       1570 ACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTG 1629
Db       2112 AGCATCGCCGCCGGTTTCGCCGAGCGTGCCCTGCTGACCCCCGACGCCACGGCGATCCAC 2053

Qy       1630 TCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCG 1689
Db       2052 GATGCCGCCGGCAGCTACAGCTACCGCCAGGTCGCCCAGCACGCCACCGCCCTGCGCCGC 1993
```

```
Qy       1690 CGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGC 1749
Db       1992 GTCCTGGAAGCGCACGGCGCGGGCCGTGGCCGGCGGGTCGCGGTGATGCTGCCGAAAAGC 1933

Qy       1750 TGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATC 1809
Db       1932 GCCGCGCAATTGGTCGCGGTGATCGGCATCCTCCAGGCCGGCGCCGCCTATGTGCCGGTG 1873

Qy       1810 GATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTC 1869
Db       1872 GACATCCGCCAGCCTCCGCTGCGGCGCCAGGCGATCCTCGCCAGCGCCGA---------A 1822

Qy       1870 GTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTG 1929
Db       1821 GTGGTCGCGCTGGTCTGCCTGGAAAGCGA--TGTCCCGGACGTCGGCTGCGCCTGCGTGG 1764

Qy       1930 CTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAG 1989
Db       1763 CCATCGACCGGCTG----GCCGCCGACAGCGCCTGGCCGCCACCGCCCGCGGCGGAGGTG 1708

Qy       1990 ACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGG 2049
Db       1707 GCGGCGGACGACCTCGCCTACGTGATCTACACCTCCGGCTCCACCGGCACGCCAAAGGGC 1648

Qy       2050 GTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAA 2109
Db       1647 GTGATGCTCAGCCATGCGGCGGTGAGCAACACGCTGCTCGACATCAACCAGCGCTACGGC 1588

Qy       2110 ATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTAT 2169
Db       1587 GTCGACGCCAACGACCGCGTCCTCGGCCTCGCCGAGCTGAGCTTCGACCTCTCGGTCTAC 1528

Qy       2170 GATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTG 2229
Db       1527 GACTTCTTCGGCGCCACCGCGGCGGGGGCCCAGGTGGTCCTCCCGGACCCGGCGCGCGGC 1468

Qy       2230 CGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCG 2289
Db       1467 AGCGATCCATCGCACTGGGCGGAACTGCTGGAACGCCACGCCATCACCCTGTGGAACTCG 1408

Qy       2290 GTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCT 2349
Db       1407 GTGCCGGCCCAAGGCCAGATGCTCATCGATTACCTGGAGAGCGAGCCGCAACGTCACCTG 1348

Qy       2350 AGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAG 2409
Db       1347 CCGGGACCGCGCTGCGTGCTCTGGTCCGGTGACTGGATTCCGGTCAGCCTGCCGACCCGA 1288

Qy       2410 CTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCG 2469
Db       1287 TGGTGGCGGCGCTGGCCGGACAGCGCGCTGTTCAGCCTGGGCGGCGCCACCGAGGCGGCG 1228

Qy       2470 ATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTAC 2529
Db       1227 ATCTGGTCGATCGAGCAGCCGATCCGCCCGCAGCACACCGAGCTGGCCAGCATCCCTTAT 1168

Qy       2530 GGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCG 2589
Db       1167 GGCCGTGCCCTGCGCGGGCAGAGCGTGGAAGTCCTGGATGCCCGCGGGCGGCGCTGCCCG 1108

Qy       2590 GTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGC 2649
Db       1107 CCGGGCGTGCGCGGCGAGATCCATATCGGCGGGGTGGGCCTGGCGCTCGGCTACGCCGGC 1048

Qy       2650 GATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTAC 2709
Db       1047 GATCCGCAGCGCACCGCCGAACGCTTCGTCCGTCACCCCGA---TGGCCGTCGCCTGTAT 991

Qy       2710 AAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAG 2769
Db        990 CGCACCGGCGACCTCGGCCGCTATCTGGCCGACGGCAGCATCGAGTTCCTCGGCCGCGAG 931

Qy       2770 GACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTC 2829
Db        930 GACGACCAGGTGAAGATTCGCGGCCACCGCATCGAACTGGCCGAGCTGGACGCCGCGCTG 871

Qy       2830 AAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGACGCGGCGAAC 2889
Db        870 TGCGCTCATCCGCAGGTCAATCTGGCGGCCACCGTGGTGCTCGGCGAGACCCACGAGCGC 811

Qy       2890 AAGCT 2894
Db        810 AGCCT 806
```

RESULT 11
US-09-252-991A-4815
; Sequence 4815, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4815
;   LENGTH: 2922
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4815

```
  Query Match             10.7%;  Score 455;  DB 4;  Length 2922;
  Best Local Similarity   50.1%;  Pred. No. 7.9e-91;
  Matches 1352;  Conservative    0;  Mismatches 1295;  Indels   50;  Gaps    7;

Qy        156 CCGTCAGAGACTCCCCGCAGAATCCATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTT 215
Db        264 CCGGCTGTCCGCCCCGGCAACGGTCGCGCTGCCGACGGCGCAGGATCGCGATCAGCCGTT 323

Qy        216 TCCTCTCACAGACATCCAAGAATCCTACTGGCTGGGCC---GGACAGGAGCGTTTACGGT 272
Db        324 CGAGCTGTCTTCCGTGCAGCAGGCCTACTGGCTGGGACGTGGCGCCGGCGAGGTGCTGGG 383

Qy        273 CCCAGCGGGATCCACGCCTATCGCGAATACGACTGTACGGATCTCGACGTGCCGAGGCT 332
Db        384 CAACGTCAGCTGCCATGCCTTCCTGGAATTCCGCGCGCGGGATGTCGACCCGCAGCGCCT 443

Qy        333 GAGCCGCGCCTTTCGGAAAGTCGTCGCGCGGCACGACATGCTTCGGGCCCACACGCTGCC 392
Db        444 GGCCGCGGCGGCGGAGTGCGTGCGTCAACGCCACCCGATGTTGCGGGCGCGCTTCCTCGA 503

Qy        393 CGACATGATGCAGGTGATCGAGCCTAAAGTCGACGCCGACATCGAGATCATCGATCTGCG 452
Db        504 CGGTCGCCAGCAGATCCTTCCGACGCCGCCGCTGTCCTGCTTCGACCTGCAGGACTGGCG 563

Qy        453 CGGGCTCGACCGGAGCACACGGGAAGCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCA 512
Db        564 CACCTTGCAGGTGGACGAGGCCGAGCGCGACTGGCAGGCGCTGCGCGACTGGCGCGCCCA 623

Qy        513 CCGCATCTATGACACCGAGCGCCCTCCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGA 572
Db        624 TGAATGCCTGGCGGTGGAGCGCGGCCAGGTGTTCCTGCTCGGGCTGGTGCGCATGCCGGG 683

Qy        573 GCGGCAAACCCGTCTCGTGCTCAGTATCGATCTCATTAACGTTGACCTAGGCAGCCTGTC 632
Db        684 CGGCGAGGATCGCCTCTGGCTGAGTCTCGACCTGCTTGCCGCCGATGTCGAAAGCCTGCG 743

Qy        633 CATCATCTTCAAGGACTGGCTCAGCTTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCT 692
Db        744 CCTGCTGCTGGCCGAACTGGGCGTTGCCTACCTGGCGCCGGAGCGCCTGGCGGAGCCGCC 803
```

```
Qy    693 GGAGCTCTCGTACCGCGATTATGTACTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCA 752
Db    804 GGCGCTGCATTTCGCCGACTA---CCTGGCGCACCGTGCGGCGCAACGCGCCGAGGCCGC 860

Qy    753 TCAACGATCGATGGATTACTGGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCT 812
Db    861 GGCGCGGGCCCGCGACTACTGGCTGGAACGCCTGCCGCGCTTGCCGGACGCGCCGGCCCT 920

Qy    813 TCCGATGAAGGCCGATCCATCTACCCTGAAGGAGATCCGCTTCCGGCACACGGAGCAATG 872
Db    921 GCCGTTGGCCTGCGCGCCGGAAAGCATCCGCCAGCCGCGCACCCGGCGCCTGGCATTCCA 980

Qy    873 GCTGCCGTCGGACTCCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCC 932
Db    981 GCTTTCCGCCGGCGAGAGCCGGCGCCTGGAGCGTCTTGCCGCGCAGCATGGCGTGACCTT 1040

Qy    933 GACGGGCGTCATCCTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCG 992
Db   1041 GTCCAGCGTGTTCGGCTGCGCCTTCGCGCTGGTCCTGGCGCGCTGGAGCGAAAGCGCGGA 1100

Qy    993 GTTTACGCTCAACATAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATAT 1052
Db   1101 ATTTCTCCTCAACGTGCCGTTGTTCGATCGGCATGCCGACGACCCGCGTATCGGCGAGGT 1160

Qy   1053 CACCGGGGACTTCACGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTT 1112
Db   1161 GATCGCCGACTTCACCACCCTGTTGCTGCTGGAGTGCCGGATGCAGGCCGGGGTGTCCTT 1220

Qy   1113 CGAACAGCGCGCTAAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGT 1172
Db   1221 CGCCGAGGCGGTGAAGAGCTTCCAGCGCAACCTCCACGGAGCCATCGACCACGCCGCATT 1280

Qy   1173 AAGCGGTATCGAGGTCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATT 1232
Db   1281 CCCCGCCCTGGAGGTGCTCCGCGAGGCGCGCCGG-----CAGGGGCCAGCCACGCTCGGC 1335

Qy   1233 GTTCCCCGTGGTGCTCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCA 1292
Db   1336 GCCGGTGGTGTTCGCCAGCAACCTGGGCGAGGAGGGCTTCGTCCCGGCGGCCTTCCGCGA 1395

Qy   1293 GAGGCTCGGAACTCCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCA 1352
Db   1396 CGCTTTCGGCGATCTCCACGACATGCTCTCGCAGACCCCGCAGGTCTGGCTCGACCACCA 1455

Qy   1353 GCTCTACGAGCACGATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCC 1412
Db   1456 GCTCTACCGGGTGGGCGACGGTATCCTGCTGGCCTGGGACAGCGTCGTCGGCCTGTTCCC 1515

Qy   1413 GCCCGACCTTCTGGACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGA 1472
Db   1516 CGAAGGCCTGCCGGAAACCATGTTCGAAGCCTACGTGGGGCTGCTCCAGCGTCTCTGCGA 1575

Qy   1473 GGAACCATGGGGTGAACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGC 1532
Db   1576 CAGCGCCTGGGAGCAGCCCGCCGATCTGCCGTTGCCCTGGGCGCAGCAGGCGCGCCGGGC 1635

Qy   1533 GAGCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGC 1592
Db   1636 CCTGCTCAACGGCCAGCCGGCATGCGCCACGGCGCGCACCCTGCATCGCGACTTCTTCCT 1695

Qy   1593 GCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTA 1652
Db   1696 TCGCGCCGCCGAGGCGCCGGATGCCGACGCGCTGCTCTATCGCGACCAACGTGTCACCCG 1755

Qy   1653 CGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACG 1712
Db   1756 CGGCGAACTGGCCGAGCGTGCGCTGCGCATCGCCGGCGGCCTGCGCGAAGCCGGGGTGCG 1815

Qy   1713 CCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCT 1772
Db   1816 CCCTGGCGACGCGGTCGAGGTCAGCCTGCCGCGCGGACCGCAGCAGGTCGCGGCGGTATT 1875

Qy   1773 CGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCG 1832
Db   1876 CGGCGTGCTCGCCGCAGGCGCCTGCTACGTGCCGCTGGACATCGACCAGCCGCCCGCACG 1935

Qy   1833 TATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGA 1892
Db   1936 GCGGCGCCTGATCGAAGAGGCCGCCGGGGTATGCCTGGCGATCACCGAGGAGGACGATCC 1995

Qy   1893 TGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGA 1952
Db   1996 GCAGGCCTTGCCGCCGCGCCTGGATGTCCAGCGCCTGCTGCGCGGCCCGGCGCTGGCCGC 2055

Qy   1953 AGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGT 2012
Db   2056 CCCCGTGCCGCTGGCGCCGCAGGCGA----------------------GTGCCTATGT 2091

Qy   2013 CATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGC 2072
Db   2092 GATCTACACCTCGGGCTCCACCGGGGTGCCCAAGGGCGTCGAGGTCAGCCACGCGGCGGC 2151

Qy   2073 CGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCT 2132
Db   2152 GATCAATACCATCGACGCGCTGCTCGACCTGCTGCGGGTGAACGCATCGGATCGCTTGCT 2211

Qy   2133 GGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGC 2192
Db   2212 GGCGGTCTCCGCGCTGGACTTCGATCTGTCGGTCTTCGACCTGTTCGGCGGCCTCGGCGC 2271

Qy   2193 GGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGA 2252
Db   2272 CGGTGCCAGCCTGGTCCTGCCGGCCCAGGAACAGGCGCGCGATGCCGCTGCCTGGGCGGA 2331

Qy   2253 GTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCT 2312
Db   2332 GGCTATCCAGCGGCATGCGGTGAGCCTGTGGAACTCGGCGCCGGCCTTGCTGGAGATGGC 2391

Qy   2313 CGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCT 2372
Db   2392 CCTCAGCCTGCCGGCGAGCCAGGCCGA------CTATCGCAGTCTGCGGGCGGTGCTGCT 2445

Qy   2373 GAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCC---GG 2429
Db   2446 GTCCGGCGACTGGGTGGCCCTGGACCTGCCCGGCCGCCTGCGCCCACGTTGTGCCGAAGG 2505

Qy   2430 CGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCC 2489
Db   2506 CTGCCGCCTGCATGTGCTGGGTGGCGCTACCGAAGCGGGCATCTGGTCGAACCTGCAGAG 2565

Qy   2490 CGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCA 2549
Db   2566 CGTCGATACGGTGCCGCCGCACTGGCGTTCGATTCCCTACGGCCGGCCATTGCCGGGACA 2625

Qy   2550 GACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACT 2609
Db   2626 GGCCTACCGGGTGGTCGACACCCACGGGCGCGACGTGCCGGACCTGGTGGTCGGCGAGCT 2685

Qy   2610 CTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAA 2669
Db   2686 GTGGATCGGCGGCGCCAGCCTGGCCCGCGGCTATCGCAACGATCCCGAACTCAGCGCCCG 2745

Qy   2670 GAGCTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCG 2729
Db   2746 GCGCTTCGT------CCACGATGCCCAGGGCCGCTGGTATCGCACCGGCGATCGCGGTCG 2799

Qy   2730 CTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCG 2789
Db   2800 CTACTGGGGCGACGGTACCCTGGAATTCCTCGGTCGGGTCGACCAGCAGGTGAAAGTGCG 2859

Qy   2790 CGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGT 2846
Db   2860 CGGCCAGCGCATCGAGTTGGGCGAGGTGGAGGCCGCGCTGTGCGCCCAGGCTGGCGT 2916
```

```
RESULT 12
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match              8.5%;  Score 358.2;  DB 3;  Length 4411529;
  Best Local Similarity   48.4%;  Pred. No. 7.2e-69;
  Matches 1252;  Conservative    0;  Mismatches 1283;  Indels   54;  Gaps    7;

Qy        286 CACGCCTATCGCGAATACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTT 345
Db    2675463 CACCTCTACGTCGAGTTCGATGGTGCCCGGGTCGATCCGGACAGGCTCCGCGCGGCGGCC 2675404

Qy        346 CGGAAAGTCGTCGCGCGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAG 405
Db    2675403 ACCCGGTTGGCGCTGCGGCACCCGATGCTGCGGGTGCAGTTCTTGCCCGACGGCACCCAG 2675344

Qy        406 GTGATCGAGCCTAAAGTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGG 465
Db    2675343 CGCATCCCGCCGGCGGCGGGATCTCGCGACTTTCCCATCAGTGTTGCCGACCTGCGTCAC 2675284

Qy        466 AGCACACGGGAAGCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGAC 525
Db    2675283 GTGGCCCCGGATGTCGTCGATCAGCGGTTGGCGGGGATCCGCGACGCCAAATCGCACCAG 2675224

Qy        526 ACCGAGCGCCCTCCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGT 585
Db    2675223 CAGCTCGACGGTGCGGTATTCGAACTTGCGTTGACGTTATTACCGGGGGAGCGCACCCGC 2675164

Qy        586 CTCGTGCTCAGTATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAG 645
Db    2675163 CTACATGTCGACCTGGACATGCAGGCCGCTGACGCGATGAGCTACCGCATCTTGCTGGCC 2675104

Qy        646 GACTGGCTCAGCTTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTAC 705
Db    2675103 GACCTGGCGGCCCTCTATGACGGCCGTGAACCGCCGGCACTGGGCTACACCTACCGGGAA 2675044

Qy        706 CGCGATTATGTACTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATG 765
Db    2675043 TACCGGCAGGCTATCGAGGCGGAGGAGACGCTGCCCCAACCGGTTCGCGACGCCGACCGG 2674984

Qy        766 GATTACTGGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCC 825
Db    2674983 GACTGGTGGGCGCAGCGCATCCCGCAGCTGCCGGATCCGCCCGCGCTGCCTACCCGGGCC 2674924

Qy        826 GATCCATCTACCCTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGAC 885
Db    2674923 GGCGGCGAACGCGACCGGCGCCGCAGCACCCGGCGC---TGGCACTGGCTAGACCCGCAG 2674867

Qy        886 TCCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATC 945
Db    2674866 ACCCGCGACGCGTTGTTCGCCCGAGCCCGGGCCCGCGGCATCACCCCGGCGATGACGCTG 2674807

Qy        946 CTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAAC 1005
Db    2674806 GCCGCGGCCTTTGCCAACGTGCTGGCGCGTTGGTCAGCGTCGTCGCGGTTCCTGCTGAAC 2674747

Qy       1006 ATAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTC 1065
Db    2674746 CTGCCGTTGTTCAGTCGCCAGGCCCTGCATCCAGACGTCGACCTGCTGGTCGGTGACTTC 2674687

Qy       1066 ACGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCT 1125
Db    2674686 ACCTCCTCGCTGTTGCTCGACGTCGATCTGACCGGTGCGCGCACGGCGGCGGCGCGGGCG 2674627

Qy       1126 AAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAG 1185
Db    2674626 CAGGCGGTGCAGGAAGCCCTGCGCAGCGCCGCAGGCCACAGTGCATACCCCGGGCTGTCT 2674567

Qy       1186 GTCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTG 1245
Db    2674566 GTGCTGCGTGACCTCAGCCGCCACCGTGGCACCCAGGTGCTGGCACCGGTGGTATTCACC 2674507

Qy       1246 CTCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACT 1305
Db    2674506 AGCGCGCTGGGGCTCGGCGACCTTTTCTGCCCGGACGTCACCGAGCA---ATTCGGCACA 2674450

Qy       1306 CCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCAC 1365
Db    2674449 CCCGGATGGATCATTTCGCAGGGGCCCCAGGTGCTGCTCGACGCCCAGGTCACCGAGTTC 2674390

Qy       1366 GATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTG 1425
Db    2674389 GACGGCGGTGTGCTGGTGAACTGGGATGTCCGCGAGGGGGTCTTTGCACCCGGCGTCATC 2674330

Qy       1426 GACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCT------CACTGAGGAACCA 1479
Db    2674329 GACGCCATGTTCACCCACCAGGTCGACGAATTGCTCCGGTTGGCCGCCGGGGACGACGCC 2674270

Qy       1480 TGGGGTGAACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCA 1539
Db    2674269 TGGGATGCGCCGAGCCCGTCCGCGCTACCCGCCGCGCAACGCGCGGTGCGCGCGGCGCTG 2674210

Qy       1540 AACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTC 1599
Db    2674209 AACGGTCGCACCGCCGCCCCCAGCACCGAGGCGCTGCACGACGGGTTTTTCCGCCAGGCC 2674150

Qy       1600 GAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAG 1659
Db    2674149 CAACAGCAGCCCGACGCGCCGGCGGTGTTCGCCAGTTCCGGCGACTTGAGCTACGCCCAA 2674090

Qy       1660 CTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAAC 1719
Db    2674089 CTGCGCGACCAGGCATCGGCGGTGGCCGCGGCGCTGCGTGCTGCGGGCCTACGAGTCGGC 2674030

Qy       1720 ACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTG 1779
Db    2674029 GACACCGTCGCGGTGCTGGGTCCGAAAACGGGCGAACAAGTGGCGGCTGTGCTGGGGATT 2673970

Qy       1780 CTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCAC 1839
Db    2673969 TTGGCCGCCGGCGGGGTCTATCTGCCGATCGGCGTCGACCAGCCCCGCGACCGCGCGGAG 2673910

Qy       1840 TACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAA 1899
Db    2673909 CGCATCCTGGCGACCGGTTCGGTCAACTTAGCGCT---------------CGTTTGCGGC 2673865

Qy       1900 CTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGAC 1959
Db    2673864 CCGCCATGCCAAGTGCGGGTGCCGGTCCCGACGCTGTTGCTGGCTGACGTGCTTGCCGCC 2673805

Qy       1960 GGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTAC 2019
Db    2673804 GCGCCGGCAGAATTCGTGCCCGGGCCTAGCGATCCCACCGCGCTCGCCTATGTGTTGTTC 2673745

Qy       2020 ACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAAC 2079
Db    2673744 ACCTCGGGCTCAACCGGGGAACCCAAGGGTGTCGAGGTGGCGCACGACGCCGCGATGAAC 2673685
```

```
Qy       2080 ACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTC 2139
Db    2673684 ACCGTGGAGACCTTCATCCGGCACTTCGAGCTAGGCGCCGCAGATCGCTGGCTTGCCCTG 2673625

Qy       2140 TCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGT 2199
Db    2673624 GCGACGCTGGAGTGCGACATGTCGGTGCTGGACATCTTCGCCGCCCTGCGCTCCGGCGGA 2673565

Qy       2200 ACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATC 2259
Db    2673564 GCGATCGTGGTGGTCGACGAAGCGCAGCGCCGCGATCCCGACGCCTGGGCCCGGCTTATC 2673505

Qy       2260 GAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAG 2319
Db    2673504 GACACTTACGAGGTCACGGCGTTGAATTTCATGCCGGGCTGGTTGGACATGCTGCTCGAA 2673445

Qy       2320 CATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGC 2379
Db    2673444 GTCGGCGGGGGCCGGCTGTCGTCG------------CTGCGAGCGGTGGCCGTCGGCGGC 2673397

Qy       2380 GACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTG 2439
Db    2673396 GACTGGGTGCGTCCCGACCTGGCCCGCCGCCTGCAGGTGCAAGCGCCGAGCGCACGGTTC 2673337

Qy       2440 ATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAAC 2499
Db    2673336 GCGGGGTTGGGTGGAGCCACCGAAACCGCGGTCCACGCAACCATTTTCGAGGTCCAGGAC 2673277

Qy       2500 GTCGACCTATCG---------TGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAG 2550
Db    2673276 GCGGCCAATCTGCCACCGGACTGGGCCTCGGTGCCATACGGCGTCCCGTTTCCCAACAAC 2673217

Qy       2551 ACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTC 2610
Db    2673216 GCCTGCCGGGTAGTGGCCGACAGCGGCGACGACTGCCCCGATTGGGTGGCCGGTGAGTTG 2673157

Qy       2611 TACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAG 2670
Db    2673156 TGGGTGTCCGGTCGCGGAATCGCCCGGGGTTACCGTGGCCGTCCCGAGCTGACCGCGGAG 2673097

Qy       2671 AGCTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGC 2730
Db    2673096 CGCTTCGTCGAGC------ATGACGGCCGCACCTGGTATCGCACCGGTGATCTGGCCCGC 2673043

Qy       2731 TACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGC 2790
Db    2673042 TACTGGCACGACGGCACCCTGGAGTTCGTCGGCCGTGCCGATCACCGCGTCAAGATCAGC 2672983

Qy       2791 GGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGC 2850
Db    2672982 GGGTACCGCGTCGAACTCGGCGAGATCGAAGCCGCGCTGCAGCGCTTGCCCGGTGTGCAC 2672923

Qy       2851 GACGCGGTG 2859
Db    2672922 GCGGCGGCG 2672914

RESULT 13
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             8.2%;  Score 349.2;  DB 3;  Length 4403765;
  Best Local Similarity   50.9%;  Pred. No. 6.8e-67;
  Matches 1073;  Conservative    0;  Mismatches  983;  Indels   54;  Gaps    8;

Qy        765 GGATTACTGGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGC 824
Db    2672280 GGACTGGTGGGCGCAGCGCATCCCGCAGCTGCCGGATCCGCCCGCGCTGCCTACCCGGGC 2672221

Qy        825 CGATCCATCTACCCTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGA 884
Db    2672220 CGGCGGCGAACGCGACCGGCGCCGCAGCACCCGGCGC---TGGCACTGGCTAGACCCGCA 2672164

Qy        885 CTCCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCAT 944
Db    2672163 GACCCGCGACGCGTTGTTCGCCCGAGCCCGGGCCCGCGGCATCACCCCGGCGATGACGCT 2672104

Qy        945 CCTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAA 1004
Db    2672103 GGCCGCGGCCTTTGCCAACGTGCTGGCGCGTTGGTCAGCGTCGTCGCGGTTCCTGCTGAA 2672044

Qy       1005 CATAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTT 1064
Db    2672043 CCTGCCGTTGTTCAGTCGCCAGGCCCTGCATCCAGACGTCGACCTGCTGGTCGGTGACTT 2671984

Qy       1065 CACGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGC 1124
Db    2671983 CACCTCCTCGCTGTTGCTCGACGTCGATCTGACCGGTGCGCGCACGGCGGCGGCGCGGGC 2671924

Qy       1125 TAAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGA 1184
Db    2671923 GCAGGCGGTGCAGGAAGCCCTG-CGCACGCCGCAGGCCACAGTGCATACCCCGGGCTGTC 2671865

Qy       1185 GGTCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGT 1244
Db    2671864 TGTGCTGCGTGACCTCAGCCGCCACCGTGGCACCCAGGTGCTGGCACCGGTGGTATTCAC 2671805

Qy       1245 GCTCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAAC 1304
Db    2671804 CAGCGCGCTGGGGCTCGGCGACCTTTTC--TGCCCGGACGTTCACCGAGCAATTCGGCAC 2671747

Qy       1305 TCCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCA 1364
Db    2671746 ACCCGGATGGATCATTTCGCAGGGGCCCCAGGTGCTGCTCGACGCCCAGGTCACCGAGTT 2671687

Qy       1365 CGATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCT 1424
Db    2671686 CGACGGCGGTGTGCTGGTGAACTGGGATGTCCGCGAGGGGGTCTTTGCACCCGGCGTCAT 2671627

Qy       1425 GGACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACT------GAGGAACC 1478
Db    2671626 CGACGCCATGTTCACCCACCAGGTCGACGAATTGCTCCGGTTGGCCGCCGGGGACGACGC 2671567

Qy       1479 ATGGGGTGAACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGC 1538
Db    2671566 CTGGGATGCGCCGAGCCCGTCCGCGCTACCCGCCGCGCAACGCGCGGTGCGCGCGGCGCT 2671507

Qy       1539 AAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGT 1598
Db    2671506 GAACGGTCGCACCGCCGCCCCCAGCACCGAGGCGCTGCACGACGGGTTTTTCCGCCAGGC 2671447

Qy       1599 CGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGA 1658
```

```
Db    2671446 CCAACAGCAGCCCGACGCGCCGGCGGTGTTCGCCAGTTCCGGCGACTTGAGCTACGCCCA 2671387

Qy       1659 GCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAA 1718
Db    2671386 ACTGCGCGACCAGGCATCGGCGGTGGCCGCGGCGCTGCGTGCTGCGGGCCTACGAGTCGG 2671327

Qy       1719 CACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGT 1778
Db    2671326 CGACACCGTCGCGGTGCTGGGTCCGAAAACGGGCGAACAAGTGGCGGCTGTGCTGGGGAT 2671267

Qy       1779 GCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCA 1838
Db    2671266 TTTGGCCGCCGGCGGGGTCTATCTGCCGATCGGCGTCGACCAGCCCCGCGACCGCGCGGA 2671207

Qy       1839 CTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAA 1898
Db    2671206 GCGCATCCTGGCGACCGGTTCGGTCAACTT---------------AGCGCTCGTTTGCGG 2671162

Qy       1899 ACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGA 1958
Db    2671161 CCCGCCATGCCAAGTGCGGGTGCCGGTCCCGACGCTGTTGCTGGCTGACGTGCTTGCCGC 2671102

Qy       1959 CGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTA 2018
Db    2671101 CGCGCCGGCAGAATTCGTGCCCGGGCCTAGCGATCCCACCGCGCTCGCCTATGTGTTGTT 2671042

Qy       2019 CACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAA 2078
Db    2671041 CACCTCGGGCTCAACCGGGGAACCCAAGGGTGTCGAGGTGGCGCACGACGCCGCGATGAA 2670982

Qy       2079 CACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCT 2138
Db    2670981 CACCGTGGAGACCTTCATCCGGCACTTCGAGCTAGGCGCCGCAGATCGCTGGCTTGCCCT 2670922

Qy       2139 CTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGG 2198
Db    2670921 GGCGACGCTGGAGTGCGACATGTCGGTGCTGGACATCTTCGCCGCCCTGCGCTCCGGCGG 2670862

Qy       2199 TACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGAT 2258
Db    2670861 AGCGATCGTGGTGGTCGACGAAGCGCAGCGCCGCGATCCCGACGCCTGGGCCCGGCTTAT 2670802

Qy       2259 CGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGA 2318
Db    2670801 CGACACTTACGAGGTCACGGCGTTGAATTTCATGCCGGGCTGGTTGGACATGCTGCTCGA 2670742

Qy       2319 GCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGG 2378
Db    2670741 AGTCGGCGGGGGCCGGCTG-----------TCGTCGCTGCGAGCGGTGGCCGTCGGCGG 2670694

Qy       2379 CGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGT 2438
Db    2670693 CGACTGGGTGCGTCCCGACCTGGCCCGCCGCCTGCAGGTGCAAGCGCCGAGCGCACGGTT 2670634

Qy       2439 GATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAA 2498
Db    2670633 CGCGGGGTTGGGTGGAGCCACCGAAACCGCGGTCCACGCAACCATTTTCGAGGTCCAGGA 2670574

Qy       2499 CGTCGAC---------CTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCA 2549
Db    2670573 CGCGGCCAATCTGCCACCGGACTGGGCCTCGGTGCCATACGGCGTCCCGTTTCCCAACAA 2670514

Qy       2550 GACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACT 2609
Db    2670513 CGCCTGCCGGGTAGTGGCCGACAGCGGCGACGACTGCCCCGATTGGGTGGCCGGTGAGTT 2670454

Qy       2610 CTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAA 2669
Db    2670453 GTGGGTGTCCGGTCGCGGAATCGCCCGGGGTTACCGTGGCCGTCCCGAGCTGACCGCGGA 2670394

Qy       2670 GAGCTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCG 2729
Db    2670393 GCGCTTCGTCGAGCATGACGGCCGCAC------CTGGTATCGCACCGGTGATCTGGCCCG 2670340

Qy       2730 CTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCG 2789
Db    2670339 CTACTGGCACGACGGCACCCTGGAGTTCGTCGGCCGTGCCGATCACCGCGTCAAGATCAG 2670280

Qy       2790 CGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACG 2849
Db    2670279 CGGGTACCGCGTCGAACTCGGCGAGATCGAAGCCGCGCTGCAGCGCTTGCCCGGTGTGCA 2670220

Qy       2850 CGACGCGGTG 2859
Db    2670219 CGCGGCGGCG 2670210
```

```
RESULT 14
US-09-252-991A-4678/c
; Sequence 4678, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 4678
;   LENGTH: 1842
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4678

  Query Match             8.2%;  Score 348.6;  DB 4;  Length 1842;
  Best Local Similarity   51.7%;  Pred. No. 1.6e-67;
  Matches  946;  Conservative    0;  Mismatches  839;  Indels   44;  Gaps    5;

Qy       1064 TCACGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCG 1123
Db       1842 TCACCACCCTGTTGCTGCTGGAGTGCCGGATGCAGGCCGGGGTGTCCTTCGCCGAGGCGG 1783

Qy       1124 CTAAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCG 1183
Db       1782 TGAAGAGCTTCCAGCGCAACCTCCACGGAGCCATCGACCACGCCGCATTCCCCGCCCTGG 1723

Qy       1184 AGGTCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGG 1243
Db       1722 AGGTGCTCCGCGAGGCGCGCCGG-----CAGGGGCCAGCCACGCTCGGCGCCGGTGGTGT 1668

Qy       1244 TGCTCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAA 1303
Db       1667 TCGCCAGCAACCTGGGCGAGGAGGGCTTCGTCCCGGCGGCCTTCCGCGACGCTTTCGGCG 1608

Qy       1304 CTCCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGC 1363
Db       1607 ATCTCCACGACATGCTCTCGCAGACCCCGCAGGTCTGGCTCGACCACCAGCTCTACCGGG 1548

Qy       1364 ACGATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTC 1423
Db       1547 TGGGCGACGGTATCCTGCTGGCCTGGGACAGCGTCGTCGGCCTGTTCCCCGAAGGCCTGC 1488

Qy       1424 TGGACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGG 1483
Db       1487 CGGAAACCATGTTCGAAGCCTACGTGGGGCTGCTCCAGCGTCTCTGCGACAGCGCCTGGG 1428

Qy       1484 GTGAACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACG 1543
Db       1427 AGCAGCCCGCCGATCTGCCGTTGCCCTGGGCGCAGCAGGCGCGCCGGGCCCTGCTCAACG 1368

Qy       1544 CGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGC 1603
```

```
Db        1367 GCCAGCCGGCATGCGCCACGGCGCGCACCCTGCATCGCGACTTCTTCCTTCGCGCCGCCG 1308

Qy        1604 AGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTT 1663
Db        1307 AGGCGCCGGATGCCGACGCGCTGCTCTATCGCGACCAACGTGTCACCCGCGGCGAACTGG 1248

Qy        1664 CGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACAT 1723
Db        1247 CCGAGCGTGCGCTGCGCATCGCCGGCGGCCTGCGCGAAGCCGGGGTGCGCCCTGGCGACG 1188

Qy        1724 TGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCG 1783
Db        1187 CGGTCGAGGTCAGCCTGCCGCGCGGACCGCAGCAGGTCGCGGCGGTATTCGGCGTGCTCG 1128

Qy        1784 AGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACC 1843
Db        1127 CCGCAGGCGCCTGCTACGTGCCGCTGGACATCGACCAGCCGCCCGCACGGCGGCGCCTGA 1068

Qy        1844 TCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGT 1903
Db        1067 TCGAAGAGGCCGCCGGGGTATGCCTGGCGATCACCGAGGAGGACGATCCGCAGGCCTTGC 1008

Qy        1904 CATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCG 1963
Db        1007 CGCCGCGCCTGGATGTCCAGCGCCTGCTGCGCGGCCCGGCGCTGGCCGCCCCCG------ 954

Qy        1964 ACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCT 2023
Db         953 -----------------TGCCGCTGGCGCCGCAGGCGAGTGCCTATGTGATCTACACCT 912

Qy        2024 CGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCA 2083
Db         911 CGGGCTCCACCGGGGTGCCCAAGGGCGTCGAGGTCAGCCACGCGGCGGCGATCAATACCA 852

Qy        2084 TCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCT 2143
Db         851 TCGACGCGCTGCTCGACCTGCTGCGGGTGAACGCATCGGATCGCTTGCTGGCGGTCTCCG 792

Qy        2144 CGGTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGA 2203
Db         791 CGCTGGACTTCGATCTGTCGGTCTTCGACCTGTTCGGCGGCCTCGGCGCCGGTGCCAGCC 732

Qy        2204 TCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAAC 2263
Db         731 TGGTCCTGCCGGCCCAGGAACAGGCGCGCGATGCCGCTGCCTGGGCGGAGGCTATCCAGC 672

Qy        2264 GAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATT 2323
Db         671 GGCATGCGGTGAGCCTGTGGAACTCGGCGCCGGCCTTGCTGGAGATGGCCCTCAGCCTGC 612

Qy        2324 TTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACT 2383
Db         611 CGGCGAGCCAGGCCGA------CTATCGCAGTCTGCGGGCGGTGCTGCTGTCCGGCGACT 558

Qy        2384 GGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCC---GGCGTGTCGGTGA 2440
Db         557 GGGTGGCCCTGGACCTGCCCGGCCGCCTGCGCCCACGTTGTGCCGAAGGCTGCCGCCTGC 498

Qy        2441 TCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACG 2500
Db         497 ATGTGCTGGGTGGCGCTACCGAAGCGGGCATCTGGTCGAACCTGCAGAGCGTCGATACGG 438

Qy        2501 TCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACG 2560
Db         437 TGCCGCCGCACTGGCGTTCGATTCCCTACGGCCGGCCATTGCCGGGACAGGCCTACCGGG 378

Qy        2561 TGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCG 2620
Db         377 TGGTCGACACCCACGGGCGCGACGTGCCGGACCTGGTGGTCGGCGAGCTGTGGATCGGCG 318

Qy        2621 GGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCG 2680
Db         317 GCGCCAGCCTGGCCCGCGGCTATCGCAACGATCCCGAACTCAGCGCCCGGCGCTTCGT-- 260

Qy        2681 TGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGCCCGCTACCTGCCCG 2740
Db         259 ----CCACGATGCCCAGGGCCGCTGGTATCGCACCGGCGATCGCGGTCGCTACTGGGGCG 204

Qy        2741 ATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCG 2800
Db         203 ACGGTACCCTGGAATTCCTCGGTCGGGTCGACCAGCAGGTGAAAGTGCGCGGCCAGCGCA 144

Qy        2801 TTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGA 2860
Db         143 TCGAGTTGGGCGAGGTGGAGGCCGCGCTGTGCGCCCAGGCTGGCGTGGATAGCGCCTGCG 84

Qy        2861 TTGTGCCCGTCGGGAACGACGCGGCGAAC 2889
Db          83 CGGCGGTGCTCGGCGGTGGCGTGGCGAGC 55
```

RESULT 15
US-09-252-991A-4764
; Sequence 4764, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4764
;   LENGTH: 2844
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4764

```
  Query Match              8.0%;  Score 340;  DB 4;  Length 2844;
  Best Local Similarity   51.9%;  Pred. No. 1.3e-65;
  Matches  917;  Conservative    0;  Mismatches  805;  Indels   44;  Gaps    5;

Qy        1127 AGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGG 1186
Db           1 AGAGCTTCCAGCGCAACCTCCACGGAGCCATCGACCACGCCGCATTCCCCGCCCTGGAGG 60

Qy        1187 TCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGC 1246
Db          61 TGCTCCGCGAGGCGCGCCGG-----CAGGGGCCAGCCACGCTCGGCGCCGGTGGTGTTCG 115

Qy        1247 TCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTC 1306
Db         116 CCAGCAACCTGGGCGAGGAGGGCTTCGTCCCGGCGGCCTTCCGCGACGCTTTCGGCGATC 175

Qy        1307 CGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACG 1366
Db         176 TCCACGACATGCTCTCGCAGACCCCGCAGGTCTGGCTCGACCACCAGCTCTACCGGGTGG 235

Qy        1367 ATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGG 1426
Db         236 GCGACGGTATCCTGCTGGCCTGGGACAGCGTCGTCGGCCTGTTCCCCGAAGGCCTGCCGG 295

Qy        1427 ACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTG 1486
Db         296 AAACCATGTTCGAAGCCTACGTGGGGCTGCTCCAGCGTCTCTGCGACAGCGCCTGGGAGC 355

Qy        1487 AACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGA 1546
Db         356 AGCCCGCCGATCTGCCGTTGCCCTGGGCGCAGCAGGCGCGCCGGGCCCTGCTCAACGGCC 415
```

```
Qy       1547 CCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGC 1606
Db        416 AGCCGGCATGCGCCACGGCGCGCACCCTGCATCGCGACTTCTTCCTTCGCGCCGCCGAGG 475

Qy       1607 TGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGC 1666
Db        476 CGCCGGATGCCGACGCGCTGCTCTATCGCGACCAACGTGTCACCCGCGGCGAACTGGCCG 535

Qy       1667 GCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGG 1726
Db        536 AGCGTGCGCTGCGCATCGCCGGCGGCCTGCGCGAAGCCGGGGTGCGCCCTGGCGACGCGG 595

Qy       1727 TCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGT 1786
Db        596 TCGAGGTCAGCCTGCCGCGCGGACCGCAGCAGGTCGCGGCGGTATTCGGCGTGCTCGCCG 655

Qy       1787 CAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCC 1846
Db        656 CAGGCGCCTGCTACGTGCCGCTGGACATCGACCAGCCGCCCGCACGGCGGCGCCTGATCG 715

Qy       1847 TCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCAT 1906
Db        716 AAGAGGCCGCCGGGGTATGCCTGGCGATCACCGAGGAGGACGATCCGCAGGCCTTGCCGC 775

Qy       1907 GGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACC 1966
Db        776 CGCGCCTGGATGTCCAGCGCCTGCTGCGCGGCCCGGCGCTGGCCGCCCCG--------- 826

Qy       1967 AGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGG 2026
Db        827 ---------------TGCCGCTGGCGCCGCAGGCGAGTGCCTATGTGATCTACACCTCGG 871

Qy       2027 GATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCC 2086
Db        872 GCTCCACCGGGGTGCCCAAGGGCGTCGAGGTCAGCCACGCGGCGGCGATCAATACCATCG 931

Qy       2087 TGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGC 2146
Db        932 ACGCGCTGCTCGACCTGCTGCGGGTGAACGCATCGGATCGCTTGCTGGCGGTCTCCGCGC 991

Qy       2147 TGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCG 2206
Db        992 TGGACTTCGATCTGTCGGTCTTCGACCTGTTCGGCGGCCTCGGCGCCGGTGCCAGCCTGG 1051

Qy       2207 TGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAG 2266
Db       1052 TCCTGCCGGCCCAGGAACAGGCGCGCGATGCCGCTGCCTGGGCGGAGGCTATCCAGCGGC 1111

Qy       2267 AGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTG 2326
Db       1112 ATGCGGTGAGCCTGTGGAACTCGGCGCCGGCCTTGCTGGAGATGGCCCTCAGCCTGCCGG 1171

Qy       2327 AGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGA 2386
Db       1172 CGAGCCAGGCCGA------CTATCGCAGTCTGCGGGCGGTGCTGCTGTCCGGCGACTGGG 1225

Qy       2387 TCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCC---GGCGTGTCGGTGATCA 2443
Db       1226 TGGCCCTGGACCTGCCCGGCCGCCTGCGCCCACGTTGTGCCGAAGGCTGCCGCCTGCATG 1285

Qy       2444 GCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCG 2503
Db       1286 TGCTGGGTGGCGCTACCGAAGCGGGCATCTGGTCGAACCTGCAGAGCGTCGATACGGTGC 1345

Qy       2504 ACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGC 2563
Db       1346 CGCCGCACTGGCGTTCGATTCCCTACGGCCGGCCATTGCCGGGACAGGCCTACCGGGTGG 1405

Qy       2564 TCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGG 2623
Db       1406 TCGACACCCACGGGCGCGACGTGCCGGACCTGGTGGTCGGCGAGCTGTGGATCGGCGGCG 1465

Qy       2624 TCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGC 2683
Db       1466 CCAGCCTGGCCCGCGGCTATCGCAACGATCCCGAACTCAGCGCCCGGCGCTTCGT----- 1520

Qy       2684 ACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATG 2743
Db       1521 -CCACGATGCCCAGGGCCGCTGGTATCGCACCGGCGATCGCGGTCGCTACTGGGGCGACG 1579

Qy       2744 GAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTG 2803
Db       1580 GTACCCTGGAATTCCTCGGTCGGGTCGACCAGCAGGTGAAAGTGCGCGGCCAGCGCATCG 1639

Qy       2804 AGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTG 2863
Db       1640 AGTTGGGCGAGGTGGAGGCCGCGCTGTGCGCCCAGGCTGGCGTGGATAGCGCCTGCGCGG 1699

Qy       2864 TGCCCGTCGGGAACGACGCGGCGAAC 2889
Db       1700 CGGTGCTCGGCGGTGGCGTGGCGAGC 1725
```

Search completed: October 4, 2003, 03:40:15
Job time : 267.251 secs

```
OM nucleic - nucleic search, using sw model

Run on:         October  3, 2003, 11:29:26 ; Search time 1035.8 Seconds
                                             (without alignments)
                                             11031.749 Million cell updates/sec

Title:          US-10-014-717-1_COPY_11872_16104
Perfect score:  4233
Sequence:       1 atgacgatcaatcagcttct..........agggcaggagacgtagctaa 4233

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters:      5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_19Jun03:*
                 1:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                 2:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                 3:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                 4:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                 5:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                 6:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                 7:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                 8:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                 9:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                 10:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                 11:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                 12:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                 13:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                 14:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                 15:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                 16:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                 17:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                 18:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                 19:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
                 20:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                 21:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                 22:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                 23:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                 24:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
                 25:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                    %
Result             Query
  No.    Score     Match Length DB  ID                      Description
  -----------------------------------------------------------------------
     1    4233     100.0  68750  21  AAZ55887              Sorangium cellulos
     2  4114.6      97.2  71989  21  AAA29349              Sorangium cellulos
     3     728      17.2  58857  21  AAA58471              Nucleotide sequenc
     4   543.4      12.8   4305  22  ABA89198              Escherichia coli p
c    5   543.4      12.8  48715  22  ABA89188              Escherichia coli p
c    6   543.4      12.8  50538  24  ABS78890              E. coli CFT073 gen
     7   390.4       9.2  32160  24  ABS78887              E. coli CFT073 gen
     8   360.6       8.5 135638  25  ABX34289              S. atroolivaceus l
c    9   358.2       8.5 4411529 22  AAI99682              Mycobacterium tube
c   10   349.2       8.2 4403765 22  AAI99683              Mycobacterium tube
    11   305.6       7.2   1416  23  AAS53964              Klebsiella pneumon
    12   284.2       6.7   7335  22  AAF81367              Quorum sensing con
    13     265       6.3  37856  21  AAA11992              S. cellulosum DNA
    14   261.4       6.2  13029  23  AAS51470              Pseudomonas aerugi
    15   247.6       5.8  90600  24  ABQ78872              S. roseosporus dap
    16     245       5.8   7110  22  AAF81361              Quorum sensing con
    17   241.6       5.7  77536  21  AAA14651              Nucleotide sequenc
    18     239       5.6  88421  24  AAL40781              88421nt genomic DN
    19   221.8       5.2   7347  23  AAS54136              Pseudomonas aerugi
c   20   220.6       5.2   5305  22  ABA89270              Escherichia coli p
    21   211.2       5.0   1566  22  AAF26321              Pseudomonas sp lip
c   22   203.8       4.8 135638  25  ABX34289              S. atroolivaceus l
    23   202.4       4.8   5451  22  AAF26319              Pseudomonas sp lip
    24   202.2       4.8   1706  22  ABA89271              Escherichia coli p
    25     188       4.4   3348  22  ABA89272              Escherichia coli p
    26   181.4       4.3   1341  22  AAF26322              Pseudomonas sp lip
    27   179.4       4.2  18660  21  AAA58472              Nucleotide sequenc
    28   177.4       4.2   4341  25  ABT32151              Benzodiazepines bi
    29   177.4       4.2  32539  25  ABT32129              Benzodiazepines bi
    30   172.6       4.1   7178  21  AAA59145              DNA encoding a pep
    31   172.2       4.1   3849  22  AAF25795              S. chrysomallus ac
    32   171.6       4.1  23666  12  AAQ10190              Cephalosporin anti
    33   168.4       4.0   8301  22  AAF90035              Nucleotide sequenc
    34   168.4       4.0  34071  22  AAF90033              Nucleotide sequenc
c   35   168.4       4.0  42717  22  AAF90032              Nucleotide sequenc
c   36   158.2       3.7   1178  20  AAV69562              Soil derived pepti
    37   155.2       3.7   3471  24  ABK74876              Bacillus lichenifo
    38     155       3.7   1200  20  AAV69559              Soil derived pepti
    39   152.6       3.6   6390  22  AAF81370              Quorum sensing con
c   40   150.8       3.6   1204  20  AAV69560              Soil derived pepti
    41   148.4       3.5   1803  25  ABT32150              Benzodiazepines bi
    42   138.6       3.3  31122  14  AAQ40706              Bacillus subtilis
c   43   137.4       3.2   1172  20  AAV69527              Soil bacteria pept
    44   137.4       3.2   4615  22  AAF90034              Nucleotide sequenc
    45   136.6       3.2   3798  24  ABK74884              Bacillus lichenifo

                                  ALIGNMENTS

RESULT 1
AAZ55887
ID   AAZ55887 standard; DNA; 68750 BP.
XX
AC   AAZ55887;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum 68.75 kb contig.
XX
KW   Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW   anticancer; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1826
FT                   /*tag=  a
FT                   /partial
FT                   /product= "Partial Orf 1 protein (AAY58580)"
FT                   /note= "No initiation codon given in the specification"
FT   CDS             complement (1900..3171)
FT                   /*tag=  b
FT                   /product= "Orf 2 protein (AAY58581)"
FT   CDS             3415..5556
FT                   /*tag=  c
FT                   /product= "Orf 3 protein (AAY58582)"
FT   CDS             complement (5612..5992)
FT                   /*tag=  d
FT                   /product= "Orf 4 protein (AAY58583)"
FT   CDS             6226..6675
```

```
FT                   /*tag=  e
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag=  f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                             (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag=  g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                             (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag=  h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                             (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag=  i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                             (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag=  j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                             (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag=  k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                             (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag=  l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                             (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag=  m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag=  n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag=  o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag=  p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag=  q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag=  r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag=  s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag=  t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag=  u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag=  v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
```

```
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR   AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR   AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR   AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy  -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             100.0%;  Score 4233;  DB 21;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy          421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy          481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy          541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy          601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy          661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy          721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy          781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy          841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy          901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy          961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy         1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy         1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy         1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy         1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy         1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy         1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy         1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy         1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy         1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy         1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy         1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy         1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy         1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy         1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy         1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy         1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy         1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy         2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy         2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy         2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy         2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy         2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy         2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy         2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy         2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy         2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451
```

```
Qy       2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy       2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy       2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy       2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy       2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy       2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy       2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy       3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy       3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy       3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy       3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy       3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy       3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy       3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||||||||||||||||||||||||||||||||
Db      16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

```
RESULT 2
AAA29349
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..992
FT                   /*tag=  a
FT                   /label= ORF_A
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             989..1501
FT                   /*tag=  b
FT                   /label= ORF_B
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             1998..6263
FT                   /*tag=  c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031..3548
```

```
FT                   /*tag=  d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                          domain"
FT   misc_RNA        3621..4661
FT                   /*tag=  e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                          domain"
FT   misc_RNA        4917..5810
FT                   /*tag=  f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                          domain, potentially involved in formation of the
FT                          thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag=  g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                          domain"
FT   CDS             6260..10493
FT                   /*tag=  h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag=  i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                          module"
FT   misc_RNA        2031..3548
FT                   /*tag=  j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                          module"
FT   misc_RNA        6861..6887
FT                   /*tag=  k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag=  l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                          module"
FT   misc_RNA        7358..7366
FT                   /*tag=  m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                          NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag=  n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag=  o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag=  p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag=  q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag=  r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag=  s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag=  t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag=  u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag=  v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                          module"
FT   misc_RNA        10261..10306
FT                   /*tag=  w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                          module"
FT   CDS             10639..16137
FT                   /*tag=  x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag=  y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag=  z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag=  aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                          module 2"
FT   misc_RNA        14962..15756
FT                   /*tag=  ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                          module 2"
FT   misc_RNA        15763..16008
FT                   /*tag=  ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag=  ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag=  ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag=  af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag=  ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag=  ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag=  ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag=  aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag=  ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag=  al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag=  am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag=  an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag=  ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag=  ap
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag=  aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag=  ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag=  as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag=  at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag=  au
```

```
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag=  av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag=  aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag=  ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag=  ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag=  az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag=  ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag=  bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             97.2%;  Score 4114.6;  DB 21;  Length 71989;
  Best Local Similarity   98.3%;  Pred. No. 0;
  Matches 4159;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       6260 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTGTCAAGCTGGCGGCCGAT 6319

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       6320 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 6379

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       6380 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAGTCC 6439

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||||
Db       6440 ATCGTGCCCGCCCCAGCCGAGCGGCACGTTCCGTTTCCTCTCACAGACATCCAAGGATCC 6499

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       6500 TACTGGCTGGGTCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 6559

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       6560 TACGACTGTACGGATCTCGACGTGGCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 6619

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6620 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 6679

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6680 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 6739

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6740 AGGCTCGTATCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 6799

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db       6800 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCAGCAAACCCGTCTCGTGCTCAGTATC 6859

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       6860 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGATTGGCTCAGCTTC 6919

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |||
Db       6920 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGACTATGTGCTC 6979

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6980 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 7039

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
              ||| ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       7040 CGCGTCGCCGAGCTCCCACCTCCGCCGATGCTTCCGATGAAGGCCGATCCATCTACCCTG 7099

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
              | ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       7100 AGGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGAGTCGATTG 7159

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
              |||| ||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db       7160 AAGCAGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATTCTGGCTGCATTTTCC 7219

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7220 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 7279

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7280 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 7339

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7340 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 7399

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7400 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 7459

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7460 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTC 7519

Qy       1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
              ||||||||||||||||||||||||||| |||||||||||||| |||||||||||||||||
Db       7520 AACCAGCAAGTCGTTGGTGTCACCTCGCTGCAGAGGCTCGGCACTCCGGTGTACACCAGC 7579

Qy       1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7580 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 7639

Qy       1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7640 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 7699

Qy       1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
              |||||||||| ||||||||||||||||||||||||||||||| |||||||| ||||||||
Db       7700 GCGTACGTCGCTTTTCTCCGGCGGCTCACTGAGGAACCATGGAGTGAACAGATGCGCTGT 7759

Qy       1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
              ||||||||||||||||||||||||||||||||||||||||||| ||||||| ||||||||
Db       7760 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGAGACCAACTCGCTGCTG 7819

Qy       1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       7820 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCTATGCAGCTC 7879

Qy       1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7880 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 7939

Qy       1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7940 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 7999

Qy       1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
```

```
Db   8000 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 8059

Qy   1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
Db   8060 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 8119

Qy   1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
Db   8120 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 8179

Qy   1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
Db   8180 CAGCGGCTGCTCGTGAGCGATGCCGGCGTCGAAGGCGACGGCGACCAGCTTCCGATGATG 8239

Qy   1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
Db   8240 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 8299

Qy   2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
Db   8300 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 8359

Qy   2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
Db   8360 CGCTTCGAAATAGGGCCCGGAGACAGAGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 8419

Qy   2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
Db   8420 TCGGTCTACGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 8479

Qy   2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
Db   8480 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGCGTTGATCGAACGAGAGAAGGTGACGGTG 8539

Qy   2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTGAGGGTCGCCCCGAT 2340
Db   8540 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTCCGAGGGTCGCCCCGAT 8599

Qy   2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
Db   8600 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 8659

Qy   2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
Db   8660 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 8719

Qy   2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
Db   8720 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGATCCATCGTGGGCGAGC 8779

Qy   2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
Db   8780 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 8839

Qy   2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
Db   8840 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGACTGGCACTGGGC 8899

Qy   2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
Db   8900 TACTGGCGCGATGAAGAGAAGACGCGCAACAGCTTCCTCGTGCACCCCGAGACCGGGGAG 8959

Qy   2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
Db   8960 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 9019

Qy   2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
Db   9020 GGGCGGGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 9079

Qy   2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
Db   9080 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 9139

Qy   2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
Db   9140 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAAGGCACACGGAGACGCGCTGCC 9199

Qy   2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
Db   9200 GAGCAGGACGCGAGCCTCAAGACCGAGCGGGTCGACGCGAGAGCACACGCCGCCAAAGCG 9259

Qy   3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
Db   9260 GACGGATTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 9319

Qy   3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
Db   9320 GATCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCTGGTTCCGCGGGAGGCGGGGCTG 9379

Qy   3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
Db   9380 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTCGAGGCCCCGATTCCATTTGTT 9439

Qy   3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
Db   9440 GAATTCGGCCGATTCCTGAGCTGCCTGAGCAGCGTGGAGCCCGACGGCGCGGCCCTTCCC 9499

Qy   3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
Db   9500 AAATTCCGTTATCCATCGGCTGGCAGCACGTACCCGGTGCAAACCTACGCGTACGCCAAA 9559

Qy   3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
Db   9560 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 9619

Qy   3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
Db   9620 TTGCTGAAGGTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCCGCAAAACTTCGAC 9679

Qy   3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
Db   9680 GTGTTCGATGAAGCGGCGTTCGGCCTCCTGTTCGTGGGCAGGATCGATGCCATCGAGTCG 9739

Qy   3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
Db   9740 CTGTATGGATCGTTGTCACGAGAATTCTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 9799

Qy   3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
Db   9800 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGTCAATTCGATTTT 9859

Qy   3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
Db   9860 GAACAGGTTCGGCCGGTTCTCGACCTGCGGCATTCGGACGTTTACGTGCACGGCATGCTG 9919

Qy   3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
Db   9920 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 9979

Qy   3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
Db   9980 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGATCAGCACTTCGCCGATATCCTT 10039

Qy   3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
Db  10040 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 10099

Qy   3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
Db  10100 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 10159

Qy   3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
Db  10160 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 10219
```

```
Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||| ||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      10220 CTCGTTGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTTGGGCTCCAGCAGAGCTTC 10279

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db      10280 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGTCTGTTGCAGAAGAGG 10339

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10340 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 10399

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              |||||||||||||||||||||| ||||||||| |||||||||||||||||||||||||||
Db      10400 TCCGGTTTGCGCCGAGACTCGAAAGATCTAGAGCAGCGGCCGAACATGCAGGACCGAGTG 10459

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||| ||||||||||||||||||||||||||||
Db      10460 GAGGCTCGGCGCAAGGGCAGGAGACGTAGCTAA 10492
```

```
RESULT 3
AAA58471
ID   AAA58471 standard; DNA; 58857 BP.
XX
AC   AAA58471;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.
XX
KW   BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW   bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW   thiazoline; bithiazoline; microbial metabolite; sugar; ss.
XX
OS   Streptomyces verticillus.
XX
FH   Key             Location/Qualifiers
FT   CDS             223..564
FT                   /*tag=  a
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 30; encodes AAB07556"
FT   CDS             561..2309
FT                   /*tag=  b
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 29; encodes AAB07557"
FT   CDS             2767..3486
FT                   /*tag=  c
FT                   /note= "ORF 28; encodes AAB07558"
FT   CDS             3527..5593
FT                   /*tag=  d
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 27; encodes AAB07559"
FT   CDS             5806..12294
FT                   /*tag=  e
FT                   /note= "ORF 26; encodes AAB07560"
FT   CDS             12291..15491
FT                   /*tag=  f
FT                   /note= "ORF 25; encodes AAB07561"
FT   CDS             15488..21013
FT                   /*tag=  g
FT                   /note= "ORF 24; encodes AAB07562"
FT   CDS             21010..24666
FT                   /*tag=  h
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 23; encodes AAB07563"
FT   CDS             24663..32690
FT                   /*tag=  i
FT                   /note= "ORF 22; encodes AAB07564"
FT   CDS             32893..34830
FT                   /*tag=  j
FT                   /note= "ORF 21; encodes AAB07565"
FT   CDS             34827..35804
FT                   /*tag=  k
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 20; encodes AAB07566"
FT   CDS             35818..37302
FT                   /*tag=  l
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 19; encodes AAB07567"
FT   CDS             37299..39215
FT                   /*tag=  m
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 18; encodes AAB07568"
FT   CDS             39301..47181
FT                   /*tag=  n
FT                   /note= "ORF 17; encodes AAB07569"
FT   CDS             47178..49985
FT                   /*tag=  o
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 16; encodes AAB07570"
FT   CDS             49982..51001
FT                   /*tag=  p
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 15; encodes AAB07571"
FT   CDS             50998..52386
FT                   /*tag=  q
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 14; encodes AAB07572"
FT   CDS             52383..52946
FT                   /*tag=  r
FT                   /note= "ORF 13; encodes AAB07573"
FT   CDS             53018..54190
FT                   /*tag=  s
FT                   /note= "ORF 12; encodes AAB07574"
FT   CDS             54187..55824
FT                   /*tag=  t
FT                   /note= "ORF 11; encodes AAB07575"
FT   CDS             55821..56093
FT                   /*tag=  u
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 10; encodes AAB07576"
FT   CDS             56090..57586
FT                   /*tag=  v
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 9; encodes AAB07577"
FT   CDS             57583..58857
FT                   /*tag=  w
FT                   /note= "ORF 8; encodes AAB07578"
XX
PN   WO200040704-A1.
XX
PD   13-JUL-2000.
XX
PF   06-JAN-2000; 2000WO-US00445.
XX
PR   06-JAN-1999;   99US-0115435.
PR   05-FEB-1999;   99US-0118848.
PR   05-JAN-2000; 2000US-0477962.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Shen B,  Du L,  Sanchez C,  Chen M,  Edwards DJ;
XX
DR   WPI; 2000-465974/40.
DR   P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
DR   AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
DR   AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
DR   AAB07576, AAB07577, AAB07578.
XX
PT   New bleomycin gene cluster components useful for peptide and/or
PT   polyketide metabolites, especially bleomycin, production and for
PT   chemically modifying biological molecules -
XX
PS   Claim 8; Page 97-136; 162pp; English.
```

```
XX
CC   The present sequence represents the BLM (Bleomycin) gene cluster,
CC   containing open reading frames (ORFs) 8-30. The proteins encoded
CC   by the gene cluster are useful for producing peptides and/or polyketide
CC   metabolites, especially bleomycin or bleomycin analogues. They are
CC   also useful for chemically modifying biological molecules to produce
CC   branched methyl groups, and for coupling amino acids and fatty
CC   acids. They may be reacted with an apo-carrier protein and coenzyme A
CC   to produce a holo-carrier protein. The BLM gene cluster or catalytic
CC   domains can be used individually or collectively to produce
CC   thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
CC   microbial metabolites. The BLM gene cluster may also be used to produce
CC   sugars.
XX
SQ   Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 other;

  Query Match             17.2%;  Score 728;  DB 21;  Length 58857;
  Best Local Similarity   55.9%;  Pred. No. 4.2e-145;
  Matches 1545;  Conservative    0;  Mismatches 1185;  Indels   36;  Gaps    7;

Qy        182 TCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCCT 241
Db      42443 TCGTCACCGACCCCGGCGCCCGGCACGAGCCGTTCCCGCTCACCGACGTCCAGCGGGCCT 42502

Qy        242 ACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGG---ATCCACGCCTATCGCG 298
Db      42503 ACTACGTGGGACGCGAGGGCGGGTTCGCCCTCGGCGGCGTCTCCACCCACGCCTACCTGG 42562

Qy        299 AATACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCG 358
Db      42563 AGATCGAGGCCCCGCGGATCGACGTCGCACGGTTTACCGGCGCGCTGCGCGGGGTGATCG 42622

Qy        359 CGCGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTA 418
Db      42623 CCCGGCACCCCATGCTGCGCGCCGTGATCCGTCCCGACGGGCTCCAGCAGGTGCTCACCG 42682

Qy        419 AAGTCGACGC---CGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGG 475
Db      42683 ACGTCCCCCCGTACGACGTGGCCGTGCACGACCTGCGCGACCTGGACGAGCCCGCGCGGC 42742

Qy        476 AAGCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCC 535
Db      42743 AGCGCCGACGCGCCGCGCTGCGCGAGGAGATGTCCCACCAGGTGGTGCCCGCCGACCTCT 42802

Qy        536 CTCCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCA 595
Db      42803 GGCCCCTGTT---CGACGTCCGCGTCTCCCTCGGCCCCACGGACGCCCTCGTCCACGTGG 42859

Qy        596 GTATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCA 655
Db      42860 GGGTGGACGCGCTGATCTGCGACGCCCACAGCTTCGGCCTCGTCCTGGCCGAACTCGCGG 42919

Qy        656 GCTTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATG 715
Db      42920 CCCGTTACGCCGACCCCGCACGCCGCTTCCCGCCCCTGACGGCGGACTTCCGGGACCACG 42979

Qy        716 TACTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGA 775
Db      42980 TCCTCCATCAGGAGGCGCTCCGCGGAACCGCCGAGTACGCGGCGGCGGAGCGGTACTGGC 43039

Qy        776 AGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTA 835
Db      43040 GCGAACGCCTGCCCGAGCTGCCGCCCGGCCCCGAACTGCCCCTGGCCGTCGCGCCCGAGA 43099

Qy        836 CCCTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTC 895
Db      43100 CCCTCGGCACCCCGCGCTTCACCCGCCGCTCCGGCCGGCTGGACGCGGCCTCCTGGACGG 43159

Qy        896 GATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCAT 955
Db      43160 CGGTCAAGGACCGGGCCCGCCGCGCCGGGCTCAGCCCCTCCGGCGTACTGCTGGCGGCGT 43219

Qy        956 TTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCT 1015
Db      43220 TCGCCGAGGTGATCACCGCGTGGAGCGGCCGGCCGCGCTACTCGCTGATGCTGACGGTCT 43279

Qy       1016 TCAACCGGGTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGG 1075
Db      43280 TCGACCGCCCGCCGCTCCACCCGGACCTCGGGCGGATCGTCGGCGACTTCACCTCGCTCA 43339

Qy       1076 TCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTC 1135
Db      43340 GCCTGCTGGAGGTCGACCACAGTCGGCCGGCGACTTCACCGACAGGGCCCGCGCCCTCC 43399

Qy       1136 AAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGT-------- 1187
Db      43400 AGCGCCGCCTGTGGCAGGACCTCGACCACCTGGCGGTCGGCGGCGTGACGGTGACACGGG 43459

Qy       1188 -------CCAGCGAGAGGCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCG 1240
Db      43460 AACGGGCGCTGCGCCACGACGCCCGACCCGGTCTGCTCACACCCGTCGTCTTCACCTCCG 43519

Qy       1241 TGGTGCTCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCG 1300
Db      43520 ACCTGCCTGTCGGCGAGACCGCGGCCGAGGACGCGGACGGGGGAGAGGGATGGGCGCTCG 43579

Qy       1301 GAACTCCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACG 1360
Db      43580 GAGAGCCCGTCTACGGCGTCAGCCAGACCCCGCAGGTCCATCTCGACCATCAAGTCGCCG 43639

Qy       1361 AGCACGATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACC 1420
Db      43640 AAGACCGAGGGGAGTTGGTCTTCAACTGGGACGCCGTGGAAGACCTGTTCGCCCCGGGCG 43699

Qy       1421 TTCTGGACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAG------G 1474
Db      43700 CCCTGGACGCCATGTTCGCCGCCTACACCGCCTCGCTGACCCGCCTGGCCCGGAGCCCCG 43759

Qy       1475 AACCATGGGGTGAACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGA 1534
Db      43760 AAGCCTGGCGGCGGCCCGGCACGCCGCCGCTGCCCACCGCCCAGGCGGCCGTGCGCCGGC 43819

Qy       1535 GCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGC 1594
Db      43820 GCACCGCCGCGACCGAGGCGCCCCTGCCCGCCCGCCTGCTGCACGAGGCCGTCGGCGACG 43879

Qy       1595 GGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACG 1654
Db      43880 CGGCCCGGCGCCACGCCGACCTGACCGCCCTGGTCGACGGCGACACCCGGATGACCTACC 43939

Qy       1655 AAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCC 1714
Db      43940 GGCGACTGACCGAGCACGCCCGGCGCGTCGGCCGCACGCTGCGCCGCCTCGGCGCCCGCC 43999

Qy       1715 CGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCG 1774
Db      44000 CCGGCCGCCTGGTCCCGGTGGTCGCCCGCAAGGGGTGGCGGCAGGCCGTCGCCGCGCTGG 44059

Qy       1775 CGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTA 1834
Db      44060 GCGTCCTGGAGTCGGGGGCGGCGTACCTGCCCCTGGACCCCGAACTGCCCGCCGAACGGC 44119

Qy       1835 TCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATG 1894
Db      44120 TCGTCCACCTCGTACGGCGCGCCGAAGCCGCCCTCCTCCTCACCGAACGCGCCCTGCTGG 44179

Qy       1895 GCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAG 1954
Db      44180 ACACGCTCGCCGTCCCCGTCGGCGTCAC---CGTGCTCGCGGTGGACGACGACGCGGCCC 44236

Qy       1955 GCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCA 2014
Db      44237 TCGACGCCGACGGCGGCCCGCTGCAGAGCGTGCAGAACCTCACCGACCTGGCGTACACCA 44296

Qy       2015 TCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCG 2074
```

```
Db      44297 TCTTCACCTCGGGCTCCACCGGCGAACCCAAGGGCGTCATGATCGACCACCTCGGCGCGG 44356

Qy       2075 TCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGG 2134
Db      44357 CCAACACCCTGGAATGCGTCAACCGCCGCTTCGGCACCGGCCCCGGCGACGCGGTCCTCG 44416

Qy       2135 CGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGG 2194
Db      44417 CCGTCTCCTCCCCGAGCTTCGACCTCGCCGTCTACGACCTGTTCGGCGTGCTGGCCGCCG 44476

Qy       2195 GCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGT 2254
Db      44477 GCGGCACCGTGGTCGTCCCCGCCCACGACCGCCGGCGCGACCCCGGACACTGGGCCGAGC 44536

Qy       2255 TGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCG 2314
Db      44537 TGATCCGGCGCGAGCGGGTCACCCTGTGGAACTCCGTCCCCGCGCTGGGCACCCTGCTCA 44596

Qy       2315 TCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGA 2374
Db      44597 CCGAGTACGCCGAGGCCCTCGCCCCCGACGCCCTGCGCACCCTGCGGGCGGTGCTCCTCA 44656

Qy       2375 GCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCAT---CAGGCCCGGCG 2431
Db      44657 GCGGCGACTGGATCCCCCTCGGACTGCCCGACCGGATCCGCGCCCTGTCCGCCCCCGGCG 44716

Qy       2432 TGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCG 2491
Db      44717 CCACCGTGATGAGCCTCGGCGGCGCGACCGAAGCCTCCATCTGGTCGGTCTGGTACGAGA 44776

Qy       2492 TGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGA 2551
Db      44777 TCGGGAAGGTGCACGAGGCGTGGAGCAGCATCCCCTACGGCACCCCCATGGCCAACCAGC 44836

Qy       2552 CGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCT 2611
Db      44837 GGCTGGAGGTCCTCGACGAGCAGCTGCGGCCCCGGCCCGACTGGGTGCCCGGCGAGCTGT 44896

Qy       2612 ACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGA 2671
Db      44897 ACATCGGCGGCACCGGCGTCGCCAAGGGCTACTGGCGCGACCCGGAACAGACCTCCCTGC 44956

Qy       2672 GCTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCT 2731
Db      44957 GCTTCCCCGTCCACCCGGGCAGCGGGCAACGCCTGTACCGCACCGGGGACTTCGCCCGCC 45016

Qy       2732 ACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCG 2791
Db      45017 ACCTCCCCGACGGCACGCTGGAATTCCTGGGCCGGCAGGACGACCAGGTGAAGATCGGCG 45076

Qy       2792 GATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCG 2851
Db      45077 GATTCCGGGTCGAACTGGGCGAGGTCGAGGCGGCCCTCGGCCGACTGCCCGACGTCGCCG 45136

Qy       2852 ACGCGGTGATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGG 2911
Db      45137 CCGGCGCGGTGATCGCCACCGGTGACCCGCGGGGCGACCGCCGCCTCGTCGGCTTCGCCG 45196

Qy       2912 TCCCGG 2917
Db      45197 TACCGG 45202
```

RESULT 4
ABA89198
ID ABA89198 standard; DNA; 4305 BP.
XX
AC ABA89198;
XX
DT 11-FEB-2002 (first entry)
XX
DE Escherichia coli polynucleotide SEQ ID NO 940.
XX
KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW systemic infection; non-diarrhoeal infection; septicaemia;
KW pyelonephritis; antibiotic resistance; ds.
XX
OS Escherichia coli.
XX
PN WO200166572-A2.
XX
PD 13-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-EP03445.
XX
PR 10-MAR-2000; 2000FR-0003145.
PR 02-FEB-2001; 2001FR-0001449.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX
DR WPI; 2001-550253/61.
XX
PT A library of DNA fragments of Escherichia coli strains for the
PT phylogenic determination of a given strain comprises polynucleotides of
PT nature B2/D+ A- -
XX
PS Example 6; Fig 6; 646pp; English.
XX
CC The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC B2/D+A-. The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicaemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more
CC frequent use of broad spectrum antibiotics.
XX
SQ Sequence 4305 BP; 880 A; 1148 C; 1309 G; 968 T; 0 other;

Query Match 12.8%; Score 543.4; DB 22; Length 4305;
Best Local Similarity 48.5%; Pred. No. 4.9e-106;
Matches 1933; Conservative 0; Mismatches 1961; Indels 93; Gaps 12;

```
Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
Db        241 ATAGTGCACGACCCTGAGCATCGTCATCAACCGTTCCCATTGACGGATGTTCAGCGCGCC 300

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
Db        301 TACTGGCTTGGACG---ACAGACGGGTGCTACCTCGATCGCGACCCACATCTACCATGAA 357

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
Db        358 TTTGACGTAGAACACTTTAATGTTACGCGTTTTACCCATGCGGTGAATGCGCTGATCGCT 417

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
Db        418 CGCCATGAAATGCTACGTGCGCGGGTACTCCCCGACGGTACTCAGCAGATTCTGGCGCAA 477

Qy        421 GTCGACGC---CGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAA 477
Db        478 GTGCCGGCGTATCAGTTAGAGCAGCGCGATCTGAGTGCTTTGTCCCCTAACGCACGAAAC 537

Qy        478 GCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCT 537
Db        538 GATGCCTTGATGGCGATCCGCGATCGGCTGTCGCATCATGTGCATCCCGCAGATCGTTGG 597

Qy        538 CCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGT 597
```

```
Db    598 CCGCTGTTTGATTTCAGTTATTCGGCTTGCACGGCGCAACATGGCCGCTTGCATTTCAGT 657
Qy    598 ATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGC 657
Db    658 CTCGATCTGCTGATTGCCGATGCTCTGAGTATGCGCACGCTACAGCAGGAGTTGATGATG 717
Qy    658 TTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTA 717
Db    718 CTGTACCGTGAGCCCCATGTGTCACTGCCGTTGCTACCGTTCTCTTTTCGTGACTACGTG 777
Qy    718 CTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAG 777
Db    778 CAGGCGCTGTTGGTAGAGCAGGCGAGTGAAGCCTATGCACGCGATCAGGCCTATTGGCAA 837
Qy    778 CGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACC 837
Db    838 CGGGCGCTGCCGCAGCTGTATGGCCCACCAACGCTGCCCGTACAGGGCGATTTGGCGCAA 897
Qy    838 CTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGA 897
Db    898 CTGTCTGCGATCAGGTTCGTACGTCGCCGTCATCGGCTGTCAGCCCACAACTGGGGAGTG 957
Qy    898 TTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTT 957
Db    958 CTGAGCGCGCTGGCCCAACGCACACGTATCACCAAGACGGCATTGTTGTTGACAGTCTTT 1017
Qy    958 TCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTC 1017
Db   1018 AGCCAAGTGCTGGCACGTTGGAGCCTTAGCCCGACGTTTACGCTCAATCTGACGTTGTTC 1077
Qy   1018 AACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTC 1077
Db   1078 AACCGCCCGCAGGGTTACCCCAACGCAGAGGCAGTTATTGGTGATTTTACCGCTGTCAGC 1137
Qy   1078 CTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAA 1137
Db   1138 TTGCTGAATGTTTGTTACGACAGCCAGCACTCTTATGCCCACAACGCTCAGCGTATTCAG 1197
Qy   1138 GAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTAT---------CGAGGTC 1188
Db   1198 GTGCAACTGTGGGAAGATCTCGAACATCGTCGTTTCAGTGGGATCCGCGCCAGCGAGGCG 1257
Qy   1189 CAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTC 1248
Db   1258 CTGATCCATAGCGGTCGTTTCCATGCGCCGATGCCGGTGGTATTCACTAGTATGTTGGAT 1317
Qy   1249 ACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCG 1308
Db   1318 ATCGACGGGGAGACGACTGCGCAAGACCCTCGGGACACAACCCGTTTTACTCTGTGTCCG 1377
Qy   1309 GTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGAT 1368
Db   1378 GACGCCAATATTACCCAAACACCGCAGGTGTGGCTCGATCACCAGGTGATCGAGTTGGCT 1437
Qy   1369 GGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGAC 1428
Db   1438 GGGGAGTTGCATTTCAACTGGGACGCGGTCGAGCAACTGTTTGATACCACGCTGCTGGAT 1497
Qy   1429 GACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAA 1488
Db   1498 CAGATGTTTGGTGCTTATTGTCATGCGCTGCAGGCGCTGGTTGCCATGCC---------G 1548
Qy   1489 CAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACC 1548
Db   1549 CAAAGTTGGTGGGGGGTAAATAGTTCTCTGGCGCTGCCCACCGTTAGTGCACCGGTCACG 1608
Qy   1549 AACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTG 1608
Db   1609 CAGGCTCCTGCACCTACGGCCTTGTTGCACCATGGATTACTGCGTCAGGCAGCACTGACG 1668
Qy   1609 CCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGC 1668
```

```
Db   1669 CCACAGGAAACTGCGCTGATCAGTCCTATCCGTGAATTGACCTATCGCCAACTGTCGACG 1728
Qy   1669 CGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTC 1728
Db   1729 GCGGCGGATCATGTGGCCCGCGCCCTGTTAGCGCTGGGCGTGCAGCATGGCGACCGCGTG 1788
Qy   1729 GCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCA 1788
Db   1789 GCGGTGGTGATGGAAAAAGGCTGGCAGCAGATTGCCGCCGTACACGGCATTTTACGACTG 1848
Qy   1789 GGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTC 1848
Db   1849 GGTGCGGTCTATCTGCCAGTGGATCCGGTGCTACCGCCACAGCGTCGCCAGCTTTTGCTG 1908
Qy   1849 GATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGG 1908
Db   1909 ACGGTGGGCGAGGTGCGGGTACAAGTAACGCAGCCGGGTCTCACGCAATTGGAGCCGTCG 1968
Qy   1909 CCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAG 1968
Db   1969 CTGCCCG------------TGCTGATCATCGACGACGGAATGCTGGACACGCCTGCTGCG 2016
Qy   1969 CCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGA 2028
Db   2017 CCGTTGCCTGAAGTGGCTGGGGATGTCACGGATCTGGCCTATATCATTTTCACTTCCGGC 2076
Qy   2029 TCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTG 2088
Db   2077 TCCACCGGTACCCCGAAAGGAGTGATGATCGACCACCGTGCGGCCATGAACACGCTGGAA 2136
Qy   2089 GACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTG 2148
Db   2137 GACATCAACGAACGCTTTGGCCTCAATGCGCAGGATAGGGTGTTCGGGCTGTCATCATTG 2196
Qy   2149 AGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTG 2208
Db   2197 AGCTTTGACCTGTCGGTTTACGATGCCTTTGCGCCTTTTATGGTGGGTGCAGCGCTGGTA 2256
Qy   2209 GTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAG 2268
Db   2257 CTGCCGGAAGCAGGACGGGAAAAAGATCCGCGTCATTGGCAGACAGTTATGGCACACGGT 2316
Qy   2269 AAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTGAG 2328
Db   2317 CATGTAAGCGTCTGGAATGCAGTGCCCGCACTGATGCAGATGCTGTGCG---AATACCAC 2373
Qy   2329 GGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATC 2388
Db   2374 AGCGGCGATCGGATGAGTTATCCGACGTTGCGTCTGGCACTGTTGAGCGGCGACTGGATC 2433
Qy   2389 CCGGTGGGCCTGCCTGGCGA---GCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGC 2445
Db   2434 CCGCTAACGTTACCGGAGCAGATGCGCGAGCGGCTCAATGAAACGATGGACATCATCAGT 2493
Qy   2446 CTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGAC 2505
Db   2494 CTGGGTGGAGCGACCGAGTGCGCCATCTGGTCGGTCTACTACCCGATAGGTGAGGTGGAA 2553
Qy   2506 CTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTC 2565
Db   2554 TCGACGTGGACCAGTATTCCCTACGGTCGGGGCCTGCGCAACCAGCCAGTATACGTGCTA 2613
Qy   2566 GATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTC 2625
Db   2614 AATGCGCAACTGGAGGAATGTCCGGTCGGGGTGGAAGGAGAGATTTGCATTGGCGGGATG 2673
Qy   2626 GGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCAC 2685
Db   2674 GGGCTGGCACAAGGCTACCTGAACGACGCAGAGAAAACGGCGGCGAGCTTTGTCTGGCGC 2733
Qy   2686 CCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGA 2745
Db   2734 GAAGCGAGTGGTGAGCGAATTTACCGCACTGGGGATCGCGGGCGCTACTTTGCTGACGGG 2793
```

```
Qy   2746 AACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAG 2805
Db   2794 CAAGTCGCCTTTTTGGGGCGCAACGATACCCAAGTGAAGGTGAATGGTTACCGTATCGAA 2853

Qy   2806 CTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTG 2865
Db   2854 CTGGGGGAAATCGAGCGCTGCATTGCGCGACATCCCGATGTGGAGCAGTCAGTGGTGGTG 2913

Qy   2866 CCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACA 2925
Db   2914 GCAGTGGGTAA---------------------------------TTCTCAACATCGTCGG 2940

Qy   2926 CGGAGACGCGCTGCCGAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCA 2985
Db   2941 CTGGTCGCTTTTGCCAAACTGCACGATCGCCACCAGGCGCAGGCATTGCAAGCTAAGGAA 3000

Qy   2986 CACGCCGCCGAAGCGGACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGA 3045
Db   3001 ------GCGGAGGCGGCGGCACTGGCGCAGGGTATTATTGTGAATCCGGCACAGCGTCTA 3054

Qy   3046 CACGGACTCCGGAGGGACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCG 3105
Db   3055 GCGTTCAAACTCAAGGAGCCACATATTCGCGCGCTGGATGGTCTGGGCATTGCACTGACG 3114

Qy   3106 CGGGAGGCGGGGCTGGACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCC 3165
Db   3115 GCACCGGCGGATAGCACACGTTACATCAAACGCCGCAGCTATCGTCATTTCAGCGCGCAA 3174

Qy   3166 CCGATTCCGTTTGTTGAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGAC 3225
Db   3175 AAAACCACGCTGGCACAGTTGGGGCAATTGCTGTCGGGCTTGGGGCAGATGCGTCTACCC 3234

Qy   3226 GGCGCGACCCTTCCCAAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACC 3285
Db   3235 GGTCTACCTTTTGCCAAGTATGCCTATGCGTCCGCCGGGGGGCTATACCCGGTGCAAACC 3294

Qy   3286 TACGCGTATGTCAAATCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCAC 3345
Db   3295 TACGTGTACCTGCATCCAGACAAGATCGAAGAGGGAGTATCCGGTATTTACTACTTCGAC 3354

Qy   3346 CCGTTCGAGCACCGTTTGCTGAAGCT---CTCCGATCACGGGATCGAGCGCGGAGCGCAC 3402
Db   3355 CCGCGACAGAGCTGTCTTATGCCGGTAGCACCAGAAGTCGAGCTGAACAGTGGTTTTCAT 3414

Qy   3403 GTTCGGCAAAACTTCGACGTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGG 3462
Db   3415 GCCGGACCTAATCAGTCTATTGCCGATCGGGCGGCATTCACGCTGTTTATGGTGGCTGAT 3474

Qy   3463 ATCGACGCCATCGAGTCGCTGTATGG---ATCGTCGTCGCGAGAATTTTGCCTGCTGGAG 3519
Db   3475 ATGGCGGTGATCTCGCCATTCTATGGGCAAGAGGCAGCTTGGCACTTCTCGGTGATGGAA 3534

Qy   3520 GCCGGATATATGGCGCAGCTCCTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGT 3579
Db   3535 GCAGGTACTCTCTGCCATTTACTGGAAGAAGATGCGCCGCGCTACGGATTGGGGCTGTGC 3594

Qy   3580 CCGGTGGGGCAATTCAATTTTGAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGAC 3639
Db   3595 CAACTTGGGATGGCAGACTTTTCCGCTGTGGCATCGCATTTTCAATTGTCGCCACATCAT 3654

Qy   3640 GTTTACGTGCACGGCATGCTGGGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACG 3699
Db   3655 CGCTATGTCCATTGCACCGTGGGGGGCGCGATAGGGCAAGAGGCGGCAAGTGCTGCAGCA 3714

Qy   3700 CTCGGTCAGGATTCCTCACCGAGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAG 3759
Db   3715 TTGCTGCGCGATTTCTCCACCTATGAGAAACCGAAGGAAACCGCTGCGCCGCTGGACATG 3774

Qy   3760 CAGCACTTCGCCGATATGCTTCGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTG 3819
Db   3775 CAGAGCTACAAAGATGCCATGCTGCGCGGCCTGCGTCAGCAACTGCCTGACTATATGGTG 3834

Qy   3820 CCTACAGTCTTCGTGGAGCTCGATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGT 3879
Db   3835 CCGAGTGATCTGATGTTAGCGACCGATTTTCCGTTAACCGCTAACGGCAAGCTAGATCGG 3894

Qy   3880 AAGGCCCTGCGCGAGCGGAAGGATACCTCGTCGCCGCGGCAT------TCGGGCACACG 3933
Db   3895 CAAAAATTACAGCTGCAGGGCGAACAAATTGCCCACCAGCGTGACGGCGTGGGTCCAATC 3954

Qy   3934 GCGCCACGGGACGCCTTGGAGGAGATCCTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTG 3993
Db   3955 CAGGTGGACAGTGCGTTACAACAGCGGCTGGTGGCGCTCTGGCAGGAGGTACTCGGCGTG 4014

Qy   3994 GAGGTGGTCGGGCTCCAGCAGAGCTTCGTCGATCTTGGTGCGACATCGATTCACATCGTT 4053
Db   4015 AGCCACGTGTCGGCCGAAGACGATTTCTTCTCGCTGGGGGGCAGTTCTATAGAATTGGTG 4074

Qy   4054 CGCATGAGGAGCCTGTTGCAGAAGAGGCTGGATAGGGAGATCGCCATCACCGAGTTGTTC 4113
Db   4075 CGTATTCAGCAGGCACTTGAGGCGATTATCGGGCAGGAGATTCCCATTGTCGATCTGTTC 4134

Qy   4114 CAGTACCCGAACCTCGGCTCGCTGGCG 4140
Db   4135 CGTCTGCCAACCATCGCGGATGTGGCG 4161
```

```
RESULT 5
ABA89188/c
ID   ABA89188 standard; DNA; 48715 BP.
XX
AC   ABA89188;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Escherichia coli polynucleotide SEQ ID NO 921.
XX
KW   Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW   immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW   systemic infection; non-diarrhoeal infection; septicaemia;
KW   pyelonephritis; antibiotic resistance; ds.
XX
OS   Escherichia coli.
XX
PN   WO200166572-A2.
XX
PD   13-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-EP03445.
XX
PR   10-MAR-2000; 2000FR-0003145.
PR   02-FEB-2001; 2001FR-0001449.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Bingen E,  Bonacorsi S,  Clermont O,  Nassif X,  Tinsley C;
XX
DR   WPI; 2001-550253/61.
XX
PT   A library of DNA fragments of Escherichia coli strains for the
PT   phylogenic determination of a given strain comprises polynucleotides of
PT   nature B2/D+ A- -
XX
PS   Example 6; Fig 6; 646pp; English.
XX
CC   The invention relates to a library of DNA fragments of Escherichia coli
CC   strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC   and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC   B2/D+A-. The polynucleotides have potential antiinflammatory,
CC   antibacterial and immunosuppressive activity as part of pharmaceutical
CC   compositions used to treat, palliate or prevent extra-intestinal E. coli
CC   infections. The polypeptides are useful for determining the phylogenic
CC   group of a given E. coli strain. These polypeptides can detect and treat
CC   an undesired development of E. coli, particularly an extra-intestinal
CC   infection that include systemic and non-diarrhoeal infections such as
```

```
CC   septicaemia, pyelonephritis and meningitis this is particularly
CC   advantageous as bacterial resistance is increasing with the more
CC   frequent use of broad spectrum antibiotics.
XX
SQ   Sequence 48715 BP; 11934 A; 13488 C; 12203 G; 11090 T; 0 other;

  Query Match             12.8%;  Score 543.4;  DB 22;  Length 48715;
  Best Local Similarity   48.5%;  Pred. No. 8.5e-106;
  Matches 1933;  Conservative    0;  Mismatches 1961;  Indels   93;  Gaps   12;

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
Db      18373 ATAGTGCACGACCCTGAGCATCGTCATCAACCGTTCCCATTGACGGATGTTCAGCGCGCC 18314

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
Db      18313 TACTGGCTTGGACG---ACAGACGGGTGCTACCTCGATCGCGACCCACATCTACCATGAA 18257

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
Db      18256 TTTGACGTAGAACACTTTAATGTTACGCGTTTTACCCATGCGGTGAATGCGCTGATCGCT 18197

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
Db      18196 CGCCATGAAATGCTACGTGCGCGGGTACTCCCCGACGGTACTCAGCAGATTCTGGCGCAA 18137

Qy        421 GTCGACGC---CGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAA 477
Db      18136 GTGCCGGCGTATCAGTTAGAGCAGCGCGATCTGAGTGCTTTGTCCCCTAACGCACGAAAC 18077

Qy        478 GCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCT 537
Db      18076 GATGCCTTGATGGCGATCCGCGATCGGCTGTCGCATCATGTGCATCCCGCAGATCGTTGG 18017

Qy        538 CCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGT 597
Db      18016 CCGCTGTTTGATTTCAGTTATTCGGCTTGCACGGCGCAACATGGCCGCTTGCATTTCAGT 17957

Qy        598 ATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGC 657
Db      17956 CTCGATCTGCTGATTGCCGATGCTCTGAGTATGCGCACGCTACAGCAGGAGTTGATGATG 17897

Qy        658 TTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTA 717
Db      17896 CTGTACCGTGAGCCCCATGTGTCACTGCCGTTGCTACCGTTCTCTTTTCGTGACTACGTG 17837

Qy        718 CTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAG 777
Db      17836 CAGGCGCTGTTGGTAGAGCAGGCGAGTGAAGCCTATGCACGCGATCAGGCCTATTGGCAA 17777

Qy        778 CGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACC 837
Db      17776 CGGGCGCTGCCGCAGCTGTATGGCCCACCAACGCTGCCCGTACAGGGCGATTTGGCGCAA 17717

Qy        838 CTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGA 897
Db      17716 CTGTCTGCGATCAGGTTCGTACGTCGCCGTCATCGGCTGTCAGCCCACAACTGGGGAGTG 17657

Qy        898 TTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTT 957
Db      17656 CTGAGCGCGCTGGCCCAACGCACACGTATCACCAAGACGGCATTGTTGTTGACAGTCTTT 17597

Qy        958 TCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTC 1017
Db      17596 AGCCAAGTGCTGGCACGTTGGAGCCTTAGCCCGACGTTTACGCTCAATCTGACGTTGTTC 17537

Qy       1018 AACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTC 1077
Db      17536 AACCGCCCGCAGGGTTACCCCAACGCAGAGGCAGTTATTGGTGATTTTACCGCTGTCAGC 17477

Qy       1078 CTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAA 1137
Db      17476 TTGCTGAATGTTTGTTACGACAGCCAGCACTCTTATGCCCACAACGCTCAGCGTATTCAG 17417

Qy       1138 GAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTAT---------CGAGGTC 1188
Db      17416 GTGCAACTGTGGGAAGATCTCGAACATCGTCGTTTCAGTGGGATCCGCGCCAGCGAGGCG 17357

Qy       1189 CAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTC 1248
Db      17356 CTGATCCATAGCGGTCGTTTCCATGCGCCGATGCCGGTGGTATTCACTAGTATGTTGGAT 17297

Qy       1249 ACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCG 1308
Db      17296 ATCGACGGGGAGACGACTGCGCAAGACCCTCGGGACACAACCCGTTTACTCTGTGTCCG 17237

Qy       1309 GTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGAT 1368
Db      17236 GACGCCAATATTACCCAAACACCGCAGGTGTGGCTCGATCACCAGGTGATCGAGTTGGCT 17177

Qy       1369 GGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGAC 1428
Db      17176 GGGGAGTTGCATTTCAACTGGGACGCGGTCGAGCAACTGTTTGATACCACGCTGCTGGAT 17117

Qy       1429 GACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAA 1488
Db      17116 CAGATGTTTGGTGCTTATTGTCATGCGCTGCAGGCGCTGGTTGCCATGCC---------G 17066

Qy       1489 CAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACC 1548
Db      17065 CAAAGTTGGTGGGGGGTAAATAGTTCTCTGGCGCTGCCCACCGTTAGTGCACCGGTCACG 17006

Qy       1549 AACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTG 1608
Db      17005 CAGGCTCCTGCACCTACGGCCTTGTTGCACCATGGATTACTGCGTCAGGCAGCACTGACG 16946

Qy       1609 CCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGC 1668
Db      16945 CCACAGGAAACTGCGCTGATCAGTCCTATCCGTGAATTGACCTATCGCCAACTGTCGACG 16886

Qy       1669 CGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTC 1728
Db      16885 GCGGCGGATCATGTGGCCCGCGCCCTGTTAGCGCTGGGCGTGCAGCATGGCGACCGCGTG 16826

Qy       1729 GCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCA 1788
Db      16825 GCGGTGGTGATGGAAAAAGGCTGGCAGCAGATTGCCGCCGTACACGGCATTTTACGACTG 16766

Qy       1789 GGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTC 1848
Db      16765 GGTGCGGTCTATCTGCCAGTGGATCCGGTGCTACCGCCACAGCGTCGCCAGCTTTTGCTG 16706

Qy       1849 GATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGG 1908
Db      16705 ACGGTGGGCGAGGTGCGGGTACAAGTAACGCAGCCGGGTCTCACGCAATTGGAGCCGTCG 16646

Qy       1909 CCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAG 1968
Db      16645 CTGCCCG------------TGCTGATCATCGACGACGGAATGCTGGACACGCCTGCTGCG 16598

Qy       1969 CCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGA 2028
Db      16597 CCGTTGCCTGAAGTGGCTGGGGATGTCACGGATCTGGCCTATATCATTTTCACTTCCGGC 16538

Qy       2029 TCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTG 2088
Db      16537 TCCACCGGTACCCCGAAAGGAGTGATGATCGACCACCGTGCGGCCATGAACACGCTGGAA 16478

Qy       2089 GACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTG 2148
Db      16477 GACATCAACGAACGCTTTGGCCTCAATGCGCAGGATAGGGTGTTCGGGCTGTCATCATTG 16418

Qy       2149 AGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTG 2208
Db      16417 AGCTTTGACCTGTCGGTTTACGATGCCTTTGCGCCTTTTATGGTGGGTGCAGCGCTGGTA 16358
```

```
Qy     2209 GTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAG 2268
Db    16357 CTGCCGGAAGCAGGACGGGAAAAAGATCCGCGTCATTGGCAGACAGTTATGGCACACGGT 16298

Qy     2269 AAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAG 2328
Db    16297 CATGTAAGCGTCTGGAATGCAGTGCCCGCACTGATGCAGATGCTGTGCG---AATACCAC 16241

Qy     2329 GGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATC 2388
Db    16240 AGCGGCGATCGGATGAGTTATCCGACGTTGCGTCTGGCACTGTTGAGCGGCGACTGGATC 16181

Qy     2389 CCGGTGGGCCTGCCTGGCGA---GCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGC 2445
Db    16180 CCGCTAACGTTACCGGAGCAGATGCGCGAGCGGCTCAATGAAACGATGGACATCATCAGT 16121

Qy     2446 CTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGAC 2505
Db    16120 CTGGGTGGAGCGACCGAGTGCGCCATCTGGTCGGTCTACTACCCGATAGGTGAGGTGGAA 16061

Qy     2506 CTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTC 2565
Db    16060 TCGACGTGGACCAGTATTCCCTACGGTCGGGGCCTGCGCAACCAGCCAGTATACGTGCTA 16001

Qy     2566 GATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTC 2625
Db    16000 AATGCGCAACTGGAGGAATGTCCGGTCGGGGTGGAAGGAGAGATTTGCATTGGCGGGATG 15941

Qy     2626 GGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCAC 2685
Db    15940 GGGCTGGCACAAGGCTACCTGAACGACGCAGAGAAAACGGCGGCGAGCTTTGTCTGGCGC 15881

Qy     2686 CCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGA 2745
Db    15880 GAAGCGAGTGGTGAGCGAATTTACCGCACTGGGGATCGCGGGCGCTACTTTGCTGACGGG 15821

Qy     2746 AACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAG 2805
Db    15820 CAAGTCGCCTTTTTGGGGCGCAACGATACCCAAGTGAAGGTGAATGGTTACCGTATCGAA 15761

Qy     2806 CTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTG 2865
Db    15760 CTGGGGGAAATCGAGCGCTGCATTGCGCGACATCCCGATGTGGAGCAGTCAGTGGTGGTG 15701

Qy     2866 CCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACA 2925
Db    15700 GCAGTGGGTAA--------------------------------TTCTCAACATCGTCGG 15674

Qy     2926 CGGAGACGCGCTGCCGAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCA 2985
Db    15673 CTGGTCGCTTTTGCCAAACTGCACGATCGCCACCAGGCGCAGGCATTGCAAGCTAAGGAA 15614

Qy     2986 CACGCCGCCGAAGCGGACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGA 3045
Db    15613 ------GCGGAGGCGGCGGCACTGGCGCAGGGTATTATTGTGAATCCGGCACAGCGTCTA 15560

Qy     3046 CACGGACTCCGGAGGGACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCG 3105
Db    15559 GCGTTCAAACTCAAGGAGCCACATATTCGCGCGCTGGATGGTCTGGGCATTGCACTGACG 15500

Qy     3106 CGGGAGGCGGGGCTGGACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCC 3165
Db    15499 GCACCGGCGGATAGCACACGTTACATCAAACGCCGCAGCTATCGTCATTTCAGCGCGCAA 15440

Qy     3166 CCGATTCCGTTTGTTGAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGAC 3225
Db    15439 AAAACCACGCTGGCACAGTTGGGGCAATTGCTGTCGGGCTTGGGGCAGATGCGTCTACCC 15380

Qy     3226 GGCGCGACCCTTCCCAAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACC 3285
Db    15379 GGTCTACCTTTTGCCAAGTATGCCTATGCGTCCGCCGGGGGCTATACCCGGTGCAAACC 15320

Qy     3286 TACGCGTATGTCAAATCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCAC 3345
Db    15319 TACGTGTACCTGCATCCAGACAAGATCGAAGAGGGAGTATCCGGTATTTACTACTTCGAC 15260

Qy     3346 CCGTTCGAGCACCGTTTGCTGAAGCT---CTCCGATCACGGGATCGAGCGCGGAGCGCAC 3402
Db    15259 CCGCGACAGAGCTGTCTTATGCCGGTAGCACCAGAAGTCGAGCTGAACAGTGGTTTTCAT 15200

Qy     3403 GTTCGGCAAAACTTCGACGTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGG 3462
Db    15199 GCCGGACCTAATCAGTCTATTGCCGATCGGGCGGCATTCACGCTGTTTATGGTGGCTGAT 15140

Qy     3463 ATCGACGCCATCGAGTCGCTGTATGG---ATCGTCGTCGCGAGAATTTTGCCTGCTGGAG 3519
Db    15139 ATGGCGGTGATCTCGCCATTCTATGGCAAGAGGCAGCTTGGCACTTCTCGGTGATGGAA 15080

Qy     3520 GCCGGATATATGGCGCAGCTCCTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGT 3579
Db    15079 GCAGGTACTCTCTGCCATTTACTGGAAGAAGATGCGCCGCGCTACGGATTGGGGCTGTGC 15020

Qy     3580 CCGGTGGGGCAATTCAATTTTGAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGAC 3639
Db    15019 CAACTTGGGATGGCAGACTTTTCCGCTGTGGCATCGCATTTCAATTGTCGCCACATCAT 14960

Qy     3640 GTTTACGTGCACGGCATGCTGGGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACG 3699
Db    14959 CGCTATGTCCATTGCACCGTGGGGGGCGCGATAGGGCAAGAGGCGGCAAGTGCTGCAGCA 14900

Qy     3700 CTCGGTCAGGATTCCTCACCGAGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAG 3759
Db    14899 TTGCTGCGCGATTTCTCCACCTATGAGAAACCGAAGGAAACCGCTGCGCCGCTGGACATG 14840

Qy     3760 CAGCACTTCGCCGATATGCTTCGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTG 3819
Db    14839 CAGAGCTACAAAGATGCCATGCTGCGCGGCCTGCGTCAGCAACTGCCTGACTATATGGTG 14780

Qy     3820 CCTACAGTCTTCGTGGAGCTCGATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGT 3879
Db    14779 CCGAGTGATCTGATGTTAGCGACCGATTTTCCGTTAACCGCTAACGGCAAGCTAGATCGG 14720

Qy     3880 AAGGCCCTGCGCGAGCGGAAGGATACCTCGTCGCCGCGGCAT------TCGGGGCACACG 3933
Db    14719 CAAAAATTACAGCTGCAGGGCGAACAAATTGCCCACCAGCGTGACGGCGTGGGTCCAATC 14660

Qy     3934 GCGCCACGGGACGCCTTGGAGGAGATCCTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTG 3993
Db    14659 CAGGTGGACAGTGCGTTACAACAGCGGCTGGTGGCGCTCTGGCAGGAGGTACTCGGCGTG 14600

Qy     3994 GAGGTGGTCGGGCTCCAGCAGAGCTTCGTCGATCTTGGTGCGACATCGATTCACATCGTT 4053
Db    14599 AGCCACGTGTCGGCCGAAGACGATTTCTTCTCGCTGGGGGGCAGTTCTATAGAATTGGTG 14540

Qy     4054 CGCATGAGGAGCCTGTTGCAGAAGAGGCTGGATAGGGAGATCGCCATCACCGAGTTGTTC 4113
Db    14539 CGTATTCAGCAGGCACTTGAGGCGATTATCGGGCAGGAGATTCCCATTGTCGATCTGTTC 14480

Qy     4114 CAGTACCCGAACCTCGGCTCGCTGGCG 4140
Db    14479 CGTCTGCCAACCATCGCGGATGTGGCG 14453
```

RESULT 6
ABS78890/c
ID ABS78890 standard; DNA; 50538 BP.
XX
AC ABS78890;
XX
DT 17-DEC-2002 (first entry)
XX
DE E. coli CFT073 genomic sequence #57.
XX
KW Pathogenic; Escherichia coli CFT073 infection; livestock; pyelonephritis;
KW urinary tract infection; open reading frame; ORF; uropathogenic;
KW antibacterial; aropathic; nephrotropic; gene; ds.

```
XX
OS   Escherichia coli.
XX
PN   WO200259320-A2.
XX
PD   01-AUG-2002.
XX
PF   19-OCT-2001; 2001WO-US46833.
XX
PR   19-OCT-2000; 2000US-242412P.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Blattner FR,  Welch RA,  Burland VD;
XX
DR   WPI; 2002-691532/74.
XX
PT   New DNA sequences of the pathogenic Escherichia coli CFT073 strain,
PT   useful for preventing or treating E. coli CFT073 infection in humans or
PT   livestock  -
XX
PS   Claim 1; Page 186-214; 765pp; English.
XX
CC   The present invention relates to polynucleotide sequences from the
CC   genome of the pathogenic Escherichia coli strain CFT073. Almost all
CC   the sequences present in E. coli CFT073 are absent in the previously
CC   sequenced laboratory strain K-12. The polynucleotide sequences of
CC   the invention are useful for preventing, diagnosing or treating
CC   E. coli CFT073 infection in humans or livestock. The polynucleotide
CC   sequences are useful for preventing urinary tract infections and
CC   pyelonephritis. Likewise, the polypeptides encoded by the different
CC   open reading frames (ORF1-5) are useful for generating a vaccine
CC   against uropathogenic E. coli strains. ABS78834-ABS79085 represent
CC   genomic sequences from E. coli strain CFT073.
XX
SQ   Sequence 50538 BP; 12312 A; 13865 C; 12701 G; 11654 T; 6 other;

  Query Match             12.8%;  Score 543.4;  DB 24;  Length 50538;
  Best Local Similarity   48.5%;  Pred. No. 8.5e-106;
  Matches 1933;  Conservative    0;  Mismatches 1961;  Indels   93;  Gaps   12;

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
Db      20201 ATAGTGCACGACCCTGAGCATCGTCATCAACCGTTCCCATTGACGGATGTTCAGCGCGCC 20142

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
Db      20141 TACTGGCTTGGACG---ACAGACGGGTGCTACCTCGATCGCGACCCACATCTACCATGAA 20085

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
Db      20084 TTTGACGTAGAACACTTTAATGTTACGCGTTTTACCCATGCGGTGAATGCGCTGATCGCT 20025

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
Db      20024 CGCCATGAAATGCTACGTGCGCGGGTACTCCCCGACGGTACTCAGCAGATTCTGGCGCAA 19965

Qy        421 GTCGACGC---CGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAA 477
Db      19964 GTGCCGGCGTATCAGTTAGAGCAGCGCGATCTGAGTGCTTTGTCCCCTAACGCACGAAAC 19905

Qy        478 GCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCT 537
Db      19904 GATGCCTTGATGGCGATCCGCGATCGGCTGTCGCATCATGTGCATCCCGCAGATCGTTGG 19845

Qy        538 CCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGT 597
Db      19844 CCGCTGTTTGATTTCAGTTATTCGGCTTGCACGGCGCAACATGGCCGCTTGCATTTCAGT 19785

Qy        598 ATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGC 657
Db      19784 CTCGATCTGCTGATTGCCGATGCTCTGAGTATGCGCACGCTACAGCAGGAGTTGATGATG 19725

Qy        658 TTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTA 717
Db      19724 CTGTACCGTGAGCCCCATGTGTCACTGCCGTTGCTACCGTTCTCTTTTCGTGACTACGTG 19665

Qy        718 CTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAG 777
Db      19664 CAGGCGCTGTTGGTAGAGCAGGCGAGTGAAGCCTATGCACGCGATCAGGCCTATTGGCAA 19605

Qy        778 CGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACC 837
Db      19604 CGGGCGCTGCCGCAGCTGTATGGCCCACCAACGCTGCCCGTACAGGGCGATTTGGCGCAA 19545

Qy        838 CTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGA 897
Db      19544 CTGTCTGCGATCAGGTTCGTACGTCGCCGTCATCGGCTGTCAGCCCACAACTGGGGAGTG 19485

Qy        898 TTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTT 957
Db      19484 CTGAGCGCGCTGGCCCAACGCACACGTATCACCAAGACGGCATTGTTGTTGACAGTCTTT 19425

Qy        958 TCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTC 1017
Db      19424 AGCCAAGTGCTGGCACGTTGGAGCCTTAGCCCGACGTTTACGCTCAATCTGACGTTGTTC 19365

Qy       1018 AACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTC 1077
Db      19364 AACCGCCCGCAGGGTTACCCCAACGCAGAGGCAGTTATTGGTGATTTTACCGCTGTCAGC 19305

Qy       1078 CTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAA 1137
Db      19304 TTGCTGAATGTTTGTTACGACAGCCAGCACTCTTATGCCCACAACGCTCAGCGTATTCAG 19245

Qy       1138 GAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTAT---------CGAGGTC 1188
Db      19244 GTGCAACTGTGGGAAGATCTCGAACATCGTCGTTTCAGTGGGATCCGCGCCAGCGAGGCG 19185

Qy       1189 CAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTC 1248
Db      19184 CTGATCCATAGCGGTCGTTTCCATGCGCCGATGCCGGTGGTATTCACTAGTATGTTGGAT 19125

Qy       1249 ACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCG 1308
Db      19124 ATCGACGGGGAGACGACTGCGCAAGACCCTCGGGACACAACCCGTTTTACTCTGTGTCCG 19065

Qy       1309 GTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGAT 1368
Db      19064 GACGCCAATATTACCCAAACACCGCAGGTGTGGCTCGATCACCAGGTGATCGAGTTGGCT 19005

Qy       1369 GGGGACCTGGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGAC 1428
Db      19004 GGGGAGTTGCATTTCAACTGGGACGCGGTCGAGCAACTGTTTGATACCACGCTGCTGGAT 18945

Qy       1429 GACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAA 1488
Db      18944 CAGATGTTTGGTGCTTATTGTCATGCGCTGCAGGCGCTGGTTGCCATGCC---------G 18894

Qy       1489 CAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACC 1548
Db      18893 CAAAGTTGGTGGGGGGTAAATAGTTCTCTGGCGCTGCCCACCGTTAGTGCACCGGTCACG 18834

Qy       1549 AACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTG 1608
Db      18833 CAGGCTCCTGCACCTACGGCCTTGTTGCACCATGGATTACTGCGTCAGGCAGCACTGACG 18774

Qy       1609 CCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGC 1668
Db      18773 CCACAGGAAACTGCGCTGATCAGTCCTATCCGTGAATTGACCTATCGCCAACTGTCGACG 18714

Qy       1669 CGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTC 1728
Db      18713 GCGGCGGATCATGTGGCCCGCGCCCTGTTAGCGCTGGGCGTGCAGCATGGCGACCGCGTG 18654

Qy       1729 GCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCA 1788
```

```
Db    18653 GCGGTGGTGATGGAAAAAGGCTGGCAGCAGATTGCCGCCGTACACGGCATTTTACGACTG 18594

Qy     1789 GGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTC 1848
Db    18593 GGTGCGGTCTATCTGCCAGTGGATCCGGTGCTACCGCCACAGCGTCGCCAGCTTTTGCTG 18534

Qy     1849 GATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGG 1908
Db    18533 ACGGTGGGCGAGGTGCGGGTACAAGTAACGCAGCCGGGTCTCACGCAATTGGAGCCGTCG 18474

Qy     1909 CCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAG 1968
Db    18473 CTGCCCG------------TGCTGATCATCGACGACGGAATGCTGGACACGCCTGCTGCG 18426

Qy     1969 CCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGA 2028
Db    18425 CCGTTGCCTGAAGTGGCTGGGGATGTCACGGATCTGGCCTATATCATTTTCACTTCCGGC 18366

Qy     2029 TCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTG 2088
Db    18365 TCCACCGGTACCCCGAAAGGAGTGATGATCGACCACCGTGCGGCCATGAACACGCTGGAA 18306

Qy     2089 GACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTG 2148
Db    18305 GACATCAACGAACGCTTTGGCCTCAATGCGCAGGATAGGGTGTTCGGGCTGTCATCATTG 18246

Qy     2149 AGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGCGGTACGATCGTG 2208
Db    18245 AGCTTTGACCTGTCGGTTTACGATGCCTTTGCGCCTTTTATGGTGGGTGCAGCGCTGGTA 18186

Qy     2209 GTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAG 2268
Db    18185 CTGCCGGAAGCAGGACGGGAAAAAGATCCGCGTCATTGGCAGACAGTTATGGCACACGGT 18126

Qy     2269 AAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAG 2328
Db    18125 CATGTAAGCGTCTGGAATGCAGTGCCCGCACTGATGCAGATGCTGTGCG---AATACCAC 18069

Qy     2329 GGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATC 2388
Db    18068 AGCGGCGATCGGATGAGTTATCCGACGTTGCGTCTGGCACTGTTGAGCGGCGACTGGATC 18009

Qy     2389 CCGGTGGGCCTGCCTGGCGA---GCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGC 2445
Db    18008 CCGCTAACGTTACCGGAGCAGATGCGCGAGCGGCTCAATGAAACGATGGACATCATCAGT 17949

Qy     2446 CTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGAC 2505
Db    17948 CTGGGTGGAGCGACCGAGTGCGCCATCTGGTCGGTCTACTACCCGATAGGTGAGGTGGAA 17889

Qy     2506 CTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTC 2565
Db    17888 TCGACGTGGACCAGTATTCCCTACGGTCGGGGCCTGCGCAACCAGCCAGTATACGTGCTA 17829

Qy     2566 GATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTC 2625
Db    17828 AATGCGCAACTGGAGGAATGTCCGGTCGGGGTGGAAGGAGAGATTTGCATTGGCGGGATG 17769

Qy     2626 GGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCAC 2685
Db    17768 GGGCTGGCACAAGGCTACCTGAACGACGCAGAGAAAACGGCGGCGAGCTTTGTCTGGCGC 17709

Qy     2686 CCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGA 2745
Db    17708 GAAGCGAGTGGTGAGCGAATTTACCGCACTGGGGATCGCGGGCGCTACTTTGCTGACGGG 17649

Qy     2746 AACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAG 2805
Db    17648 CAAGTCGCCTTTTTGGGGCGCAACGATACCCAAGTGAAGGTGAATGGTTACCGTATCGAA 17589

Qy     2806 CTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTG 2865
Db    17588 CTGGGGGAAATCGAGCGCTGCATTGCGCGACATCCCGATGTGGAGCAGTCAGTGGTGGTG 17529

Qy     2866 CCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACA 2925
Db    17528 GCAGTGGGTAA----------------------------------TTCTCAACATCGTCGG 17502

Qy     2926 CGGAGACGCGCTGCCGAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCA 2985
Db    17501 CTGGTCGCTTTTGCCAAACTGCACGATCGCCACCAGGCGCAGGCATTGCAAGCTAAGGAA 17442

Qy     2986 CACGCCGCCGAAGCGGACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGA 3045
Db    17441 ------GCGGAGGCGGCGGCACTGGCGCAGGGTATTATTGTGAATCCGGCACAGCGTCTA 17388

Qy     3046 CACGGACTCCGGAGGGACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCG 3105
Db    17387 GCGTTCAAACTCAAGGAGCCACATATTCGCGCGCTGGATGGTCTGGGCATTGCACTGACG 17328

Qy     3106 CGGGAGGCGGGGCTGGACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCC 3165
Db    17327 GCACCGGCGGATAGCACACGTTACATCAAACGCCGCAGCTATCGTCATTTCAGCGCGCAA 17268

Qy     3166 CCGATTCCGTTTGTTGAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGAC 3225
Db    17267 AAAACCACGCTGGCACAGTTGGGGCAATTGCTGTCGGGCTTGGGGCAGATGCGTCTACCC 17208

Qy     3226 GGCGCGACCCTTCCCAAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACC 3285
Db    17207 GGTCTACCTTTTGCCAAGTATGCCTATGCGTCCGCCGGGGGGCTATACCCGGTGCAAACC 17148

Qy     3286 TACGCGTATGTCAAATCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCAC 3345
Db    17147 TACGTGTACCTGCATCCAGACAAGATCGAAGAGGGAGTATCCGGTATTTACTACTTCGAC 17088

Qy     3346 CCGTTCGAGCACCGTTTGCTGAAGCT---CTCCGATCACGGGATCGAGCGCGGAGCGCAC 3402
Db    17087 CCGCGACAGAGCTGTCTTATGCCGGTAGCACCAGAAGTCGAGCTGAACAGTGGTTTTCAT 17028

Qy     3403 GTTCGGCAAAACTTCGACGTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGG 3462
Db    17027 GCCGGACCTAATCAGTCTATTGCCGATCGGGCGGCATTCACGCTGTTTATGGTGGCTGAT 16968

Qy     3463 ATCGACGCCATCGAGTCGCTGTATGG---ATCGTCGTCGCGAGAATTTTGCCTGCTGGAG 3519
Db    16967 ATGGCGGTGATCTCGCCATTCTATGGGCAAGAGGCAGCTTGGCACTTCTCGGTGATGGAA 16908

Qy     3520 GCCGGATATATGGCGCAGCTCCTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGT 3579
Db    16907 GCAGGTACTCTCTGCCATTTACTGGAAGAAGATGCGCCGCGCTACGGATTGGGGCTGTGC 16848

Qy     3580 CCGGTGGGGCAATTCAATTTTGAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGAC 3639
Db    16847 CAACTTGGGATGGCAGACTTTTCCGCTGTGGCATCGCATTTTCAATTGTCGCCACATCAT 16788

Qy     3640 GTTTACGTGCACGGCATGCTGGGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACG 3699
Db    16787 CGCTATGTCCATTGCACCGTGGGGGGCGCGATAGGGCAAGAGGCGGCAAGTGCTGCAGCA 16728

Qy     3700 CTCGGTCAGGATTCCTCACCGAGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAG 3759
Db    16727 TTGCTGCGCGATTTCTCCACCTATGAGAAACCGAAGGAAACCGCTGCGCCGCTGGACATG 16668

Qy     3760 CAGCACTTCGCCGATATGCTTCGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTG 3819
Db    16667 CAGAGCTACAAAGATGCCATGCTGCGCGGCCTGCGTCAGCAACTGCCTGACTATATGGTG 16608

Qy     3820 CCTACAGTCTTCGTGGAGCTCGATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGT 3879
Db    16607 CCGAGTGATCTGATGTTAGCGACCGATTTTCCGTTAACCGCTAACGGCAAGCTAGATCGG 16548

Qy     3880 AAGGCCCTGCGCGAGCGGAAGGATACCTCGTCGCCGCGGCAT------TCGGGGCACACG 3933
Db    16547 CAAAAATTACAGCTGCAGGGCGAACAAATTGCCCACCAGCGTGACGGCGTGGGTCCAATC 16488
```

```
Qy       3934 GCGCCACGGGACGCCTTGGAGGAGATCCTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTG 3993
Db      16487 CAGGTGGACAGTGCGTTACAACAGCGGCTGGTGGCGCTCTGGCAGGAGGTACTCGGCGTG 16428
Qy       3994 GAGGTGGTCGGGCTCCAGCAGAGCTTCGTCGATCTTGGTGCGACATCGATTCACATCGTT 4053
Db      16427 AGCCACGTGTCGGCCGAAGACGATTTCTTCTCGCTGGGGGGCAGTTCTATAGAATTGGTG 16368
Qy       4054 CGCATGAGGAGCCTGTTGCAGAAGAGGCTGGATAGGGAGATCGCCATCACCGAGTTGTTC 4113
Db      16367 CGTATTCAGCAGGCACTTGAGGCGATTATCGGGCAGGAGATTCCCATTGTCGATCTGTTC 16308
Qy       4114 CAGTACCCGAACCTCGGCTCGCTGGCG 4140
Db      16307 CGTCTGCCAACCATCGCGGATGTGGCG 16281
```

RESULT 7
ABS78887
ID   ABS78887 standard; DNA; 32160 BP.
XX
AC   ABS78887;
XX
DT   17-DEC-2002 (first entry)
XX
DE   E. coli CFT073 genomic sequence #54.
XX
KW   Pathogenic; Escherichia coli CFT073 infection; livestock; pyelonephritis;
KW   urinary tract infection; open reading frame; ORF; uropathogenic;
KW   antibacterial; aropathic; nephrotropic; gene; ds.
XX
OS   Escherichia coli.
XX
PN   WO200259320-A2.
XX
PD   01-AUG-2002.
XX
PF   19-OCT-2001; 2001WO-US46833.
XX
PR   19-OCT-2000; 2000US-242412P.
XX
PA   (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI   Blattner FR, Welch RA, Burland VD;
XX
DR   WPI; 2002-691532/74.
XX
PT   New DNA sequences of the pathogenic Escherichia coli CFT073 strain,
PT   useful for preventing or treating E. coli CFT073 infection in humans or
PT   livestock -
XX
PS   Claim 1; Page 166-184; 765pp; English.
XX
CC   The present invention relates to polynucleotide sequences from the
CC   genome of the pathogenic Escherichia coli strain CFT073. Almost all
CC   the sequences present in E. coli CFT073 are absent in the previously
CC   sequenced laboratory strain K-12. The polynucleotide sequences of
CC   the invention are useful for preventing, diagnosing or treating
CC   E. coli CFT073 infection in humans or livestock. The polynucleotide
CC   sequences are useful for preventing urinary tract infections and
CC   pyelonephritis. Likewise, the polypeptides encoded by the different
CC   open reading frames (ORF1-5) are useful for generating a vaccine
CC   against uropathogenic E. coli strains. ABS78834-ABS79085 represent
CC   genomic sequences from E. coli strain CFT073.
XX
SQ   Sequence 32160 BP; 7001 A; 9686 C; 8766 G; 6688 T; 19 other;

Query Match 9.2%; Score 390.4; DB 24; Length 32160;
Best Local Similarity 51.2%; Pred. No. 3e-73;
Matches 1086; Conservative 0; Mismatches 997; Indels 38; Gaps 6;

```
Qy        761 CGATGGATTACTGGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGA 820
Db      10021 CGCGCGCTTACTGGCTGGCGAAAGCATCGACGCTTCCCCCCGCGCCCGTCTTGCCGCTGG 10080
Qy        821 AGGCCGATCCATCTACCCTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGT 880
Db      10081 CCTGCGAACCCGCCACGCTATGTGAAGTCCGTAATACCCGACGCCGCATGATTGTTCCGG 10140
Qy        881 CGGACTCCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCG 940
Db      10141 CAACACGCTGGCACGCCTTTAGCAACCGGGCCGGCGAGTATGGCGTGACGCCGACAATGG 10200
Qy        941 TCATCCTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGC 1000
Db      10201 CGCTGGCGACCTGTTTTTCTGCCGTGCTGGCTCGCTGGGGCGGCCTGACGCGTCTGCTGC 10260
Qy       1001 TCAACATAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGG 1060
Db      10261 TTAACATCACCTTATTCGACCGCCAGCCGCTGCACCCGGCGGTTGGCGCGATGCTTGCCG 10320
Qy       1061 ACTTCACGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGC 1120
Db      10321 ACTTCACCAATATTCTTCTGCTGGACACCGCCTGCGATGGCGATACCGTCAGCAACCTGG 10380
Qy       1121 GCGCTAAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTA 1180
Db      10381 CGCGTAA---AAACCAGCTCACGTTTACGGAGGACTGGGAGCATCGCCACTGGTCCGGCG 10437
Qy       1181 TCGAGGTCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCG 1240
Db      10438 TCGAATTACTCCGTGAACTCAAACGCCAGCAGCGCTACCCCCACGGCGC------CCCGG 10491
Qy       1241 TGGTGCTCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCG 1300
Db      10492 TGGTATTTACCAGCAATCTGGGGCGTTCCCTCTACAGCAGCCGCGCAGAATCGCCGTTGG 10551
Qy       1301 GAACTCCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACG 1360
Db      10552 GCGAGCCGGAATGGGGCATCTCGCAAACGCCGCAGGTCTGGATAGATCATCTGGCGTTCG 10611
Qy       1361 AGCACGATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACC 1420
Db      10612 AGCATCACGGCGAGGTCTGGCTACAATGGGACAGCAACGACGCGCTGTTCCCTCCGGCGT 10671
Qy       1421 TTCTGGACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCAT 1480
Db      10672 TAGTCGAAACATTGTTCGACGCCTACTGCCAGTTGATTAACCAACTCTGCGATGACGAAA 10731
Qy       1481 GGGGTGAACAGGTGC--------GCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGC 1532
Db      10732 GCGCCTGGCAAAAGCCGTTCGCAGATATTATGCCCGCCAGCCAGCGCGCGATACGCGAAC 10791
Qy       1533 GAGCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGC 1592
Db      10792 GGGNTCAACKCCACCGGGTGCCCCCATTCCCGAAGGCTTGCTGCATGAAGGCATTTTCCG 10851
Qy       1593 GCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTA 1652
Db      10852 TATCGCTCTGCAACAGCCGCAGGCGCTGGCGGTAACGGACATGCGTTATCAGTGGAATTA 10911
Qy       1653 CGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACG 1712
Db      10912 TCATGAGCTGACAGACTATGCCCGCCGTTGCGCGGGCAGGTTAATCGAGTGCGGGGTTCA 10971
Qy       1713 CCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCT 1772
Db      10972 GCCCGGCGATAATGTGGCTATCACGATGTCGAAAGGCGCAGGACAACTTGTTGCGGTTCT 11031
Qy       1773 CGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCG 1832
Db      11032 GGCCGTCCTGCTGGCCGGGGCGGTTTACGTTCCGGTTTCGCTGGATCAGCCTGCCGCACG 11091
Qy       1833 TATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGA 1892
```

```
Db      11092 GCGCGAGAAAATCTACGCTGACGCCAGCGTCCGGCTGGTGCTCATTTGTCAGCACGACGC 11151

Qy       1893 TGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGA 1952

Db      11152 CAGCGCCGGGTCA-----GACGATATTCCCGCCCTTGCCTGGCAGCAGGCCATT------ 11200

Qy       1953 AGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGT 2012

Db      11201 ----GAGGCGGAGCCGATCGCCAACCCGGTAGTACGCGCCCCCACGCAACCGGCTACAT 11256

Qy       2013 CATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGC 2072

Db      11257 TATCTACACCTCCGGCTCTACCGGTACGCCGAAAGGGGTAGTCATTTCTCACCGGGGAGC 11316

Qy       2073 CGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCT 2132

Db      11317 GCTTAACACCTGTTGCGATATCAATACCCGCTATCAGGTTGGCCCGCATGACAGGGTGCT 11376

Qy       2133 GGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGC 2192

Db      11377 GGCCCTCTCCGCCCTACATTTTGATTTATCGGTTTACGACATTTTTGGCGTACTGCGCGC 11436

Qy       2193 GGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGA 2252

Db      11437 GGGCGGCGCGCTGGTGATGGTGATGGAAAATCAACGGCGCGATCCTCACGCATGGTGTGA 11496

Qy       2253 GTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCT 2312

Db      11497 GCTGATCCAGCGCCATCAGGTCACGCTCTGGAACAGCGTCCCGGCGCTGTTCGATATGCT 11556

Qy       2313 CGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCT 2372

Db      11557 GCTGACCTGGTGTGAAGGTTTCGCCGACGCCACGCCGGAAAACCTGCGCGCAGTGATGCT 11616

Qy       2373 GAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGT 2432

Db      11617 TTCCGGCGACTGGATCGGGCTTGACCTCCCCGCCCGTTATCGGGCCTTCCGGCCACAAGG 11676

Qy       2433 GTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGT 2492

Db      11677 ACAATTTATCGCGATGGGCGGCGCCACCGAGGCGTCTATCTGGTCTAACGCCTGCGAAAT 11736

Qy       2493 GAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGAC 2552

Db      11737 TCACGACGTCCCCGCCCACTGGCGCTCCATCCCTTACGGTTTTCCGCTAACCAACCAACG 11796

Qy       2553 GTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTA 2612

Db      11797 CTACCGGGTGGTGGATGAACAGGGCCGGGACTGCCCTGACTGGGTGCCGGGTGAATTATG 11856

Qy       2613 CATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAG 2672

Db      11857 GATTGGCGGCATTGGGGTCGCGGAAGGCTATTTCAACGATCCCCTGCGTAGCGAGCAGCA 11916

Qy       2673 CTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTA 2732

Db      11917 ATTTTTGACGCTCCCGGA------CGAGCGCTGGTATCGCACCGGCGATCTCGGCTGCTA 11970

Qy       2733 CCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGG 2792

Db      11971 CTGGCCAGATGGCACAATCGAGTTCCTCGGTCGTCGCGACAAGCAGGTCAAAGTCGGAGG 12030

Qy       2793 ATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGA 2852

Db      12031 ATATCGCATCGAGCTGGGCGAAATCGAAAGCGCGCTCAGCCAGCTGGCGGGGGTGAAACA 12090

Qy       2853 CGCGGTGATTGTGCCCGTCGG 2873

Db      12091 AGCAACCGTTCTGGCGATCGG 12111
```

RESULT 8
ABX34289

```
ID   ABX34289 standard; DNA; 135638 BP.
XX
AC   ABX34289;
XX
DT   11-FEB-2003  (first entry)
XX
DE   S. atroolivaceus leinamycin biosynthesis gene cluster.
XX
KW   Leinamycin biosynthesis gene cluster; Lmn; open reading frame; ORF;
KW   anti-tumour antibiotic; broad spectrum antimicrobial activity;
KW   Gram-positive; Gram-negative bacteria; chemical modification;
KW   metabolite; apo-carrier protein; holo-carrier protein; tumour;
KW   polyketide; hybrid polypeptide/polyketide metabolite; Lnm production;
KW   cytostatic; gene; ds.
XX
OS   Streptomyces atroolivaceus.
XX
PN   WO200277179-A2.
XX
PD   03-OCT-2002.
XX
PF   22-MAR-2002; 2002WO-US08937.
XX
PR   26-MAR-2001; 2001US-278935P.
XX
PA   (REGC ) UNIV CALIFORNIA.
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Shen B,  Cheng Y,  Tang G;
XX
DR   WPI; 2003-018907/01.
DR   P-PSDB; ABU11341, ABU11342, ABU11343, ABU11344, ABU11345, ABU11346,
DR   ABU11347, ABU11348, ABU11349, ABU11350, ABU11351, ABU11352, ABU11353,
DR   ABU11354, ABU11355, ABU11356, ABU11357, ABU11358, ABU11359, ABU11360,
DR   ABU11361, ABU11362, ABU11363, ABU11364, ABU11365, ABU11366, ABU11367,
DR   ABU11368, ABU11369, ABU11370, ABU11371, ABU11372, ABU11373, ABU11374,
DR   ABU11375, ABU11376, ABU11377, ABU11378, ABU11379, ABU11380, ABU11381,
DR   ABU11382, ABU11383, ABU11384, ABU11385, ABU11386, ABU11387, ABU11388,
DR   ABU11389, ABU11390, ABU11391, ABU11392, ABU11393, ABU11394, ABU11395,
DR   ABU11396, ABU11397, ABU11398, ABU11399, ABU11400, ABU11401, ABU11402,
DR   ABU11403, ABU11404, ABU11405, ABU11406, ABU11407, ABU11408, ABU11409,
DR   ABU11410, ABU11411.
XX
PT   Novel gene cluster responsible for synthesis of leinamycin in
PT   Streptomyces atroolivaceus useful for making various peptide and/or
PT   polyketide, and/or hybrid polypeptide/polyketide metabolites  -
XX
PS   Claim 6; Page 81-127; 185pp; English.
XX
CC   The present invention relates to the isolation of the Streptomyces
CC   atroolivaceus leinamycin (Lnm) biosynthesis gene cluster containing
CC   71 open reading frames (ORFs) (ORFs -35 through -1, ORFs lnmA through
CC   lnmZ, and ORFs +1 through +9). Leinamycin is a novel anti-tumour
CC   antibiotic produced by several Streptomyces species. It exhibits
CC   broad spectrum antimicrobial activity against Gram-positive and
CC   Gram-negative bacteria, but not against fungi. The polypeptides encoded
CC   by the Lnm biosynthesis gene cluster ORFs are useful for chemically
CC   modifying a molecule in a host cell. The host cell is a bacterium or
CC   eukaryotic cell, including a mammalian, yeast, plant, fungal, or insect
CC   cell. The molecule is an endogenous metabolite produced by the host
CC   cell or exogenously supplied metabolite, or an amino acid, and the
CC   polypeptide is a peptide synthetase or amino transferase. The
CC   polypeptides encoded by the Lmn gene cluster are useful for converting
CC   an apo-carrier protein to a holo-carrier protein. Lnm shows potent
CC   antitumour activity in tumour models in vivo. The Lnm gene cluster
CC   modules and/or catalytic domains are useful for making various peptide
CC   and/or polyketide, and/or hybrid polypeptide/polyketide metabolites.
CC   The proteins encoded by the ORFs are useful alone, or in combination
CC   with other active domains to modify various target substrates. The
CC   Lmn gene cluster is useful to upregulate endogenous Lnm production to
CC   permit Lnm production in cells and/or to make various modified Lnm.
CC   Lnm, its analogue, or other polyketide, peptide or hybrid
CC   polyketide/peptide metabolites are useful as therapeutic agents, to
```

```
CC   treat a number of disorders, depending upon the type of metabolites.
CC   The present sequence represents the S. atroolivaceus leinamycin
CC   biosynthesis gene cluster.
XX
SQ   Sequence 135638 BP; 18570 A; 49096 C; 49039 G; 18933 T; 0 other;

  Query Match             8.5%;  Score 360.6;  DB 25;  Length 135638;
  Best Local Similarity   55.4%;  Pred. No. 9.3e-67;
  Matches  768;  Conservative    0;  Mismatches  604;  Indels   15;  Gaps    3;

Qy       1541 ACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCG 1600
Db      72583 ACACCACCCGCGCGTTCCCCGACGACCGGCTGATGCACCAGCTCTTCGAGGAGCAGGCGG 72642

Qy       1601 AGCAGCTGCCCATGCAGCTCGCCGTGGTGTCG---GCGCGCAAGACGCTCACGTACGAAG 1657
Db      72643 CCCAGCAGCCTCGGCGGAGGCGCTGCGCTGGCGCGGCGGCGGCACGATGACATACCAGG 72702

Qy       1658 AGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGA 1717
Db      72703 AGCTGAACCGGCGGGCCAACCGCATCGCCGCCCGGCTGGCCGCCGAGGACGTGGGCCCGG 72762

Qy       1718 ACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGG 1777
Db      72763 AGACGGTCGTCGCCGTCAGTGTGCCGCGCGGGCCCATGATGGTCGCGGTGGTCCTCGGGA 72822

Qy       1778 TGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCC 1837
Db      72823 TCCTGAAGGCCGGCGGTGTCTATCTGCCGATGGAGCCGCATCTGCCCGCCGAGCGGGCCG 72882

Qy       1838 ACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCA 1897
Db      72883 CCGTCATCCTCGAGGAGGCGCACGCCGAGGTCGTCGTCACCACCGCCGACCGCGAGGGCT 72942

Qy       1898 AAC---TGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAG 1954
Db      72943 GGCCGGTGCCCGACGGATACGCCCGCGTGTGCGCGGACGCCGCCGTGGAGGGCCCCCACC 73002

Qy       1955 GCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCA 2014
Db      73003 CCGCCGACGCGGACAACTGCCCGCGGCCCGTGACGCAGCCGCACAACACGGCGTACATCA 73062

Qy       2015 TCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCG 2074
Db      73063 TCTTCACCTCGGGCAGCACCGGCCGCCCCAAGGGTGTGGCGGTCGCGCACCGGCCCGTCC 73122

Qy       2075 TCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGCCCGGAGACAGGGTGCTGG 2134
Db      73123 TGAACCTCATCAACTGGTGCCGGCGCACCTTCGGGTTCGGGCCCGGGGACATGGGCCTGT 73182

Qy       2135 CGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGG 2194
Db      73183 GTGTCACCTCGCTCGGCTTCGACCTGTCCGTGTTCGACGTGTTCGGCCTGCTCGGTACGG 73242

Qy       2195 GCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGT 2254
Db      73243 GTGCCGCGCTCTACATCGCGGACGCGGAGCAGCAGCGCGATCCCGCTCTCCTCCTGGACG 73302

Qy       2255 TGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCG 2314
Db      73303 TGCTGATCGAGGAGCCGGTCACCTTCTGGAACTCGGCGCCCACCACGCTCGCCCAGGTAG 73362

Qy       2315 TCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGA 2374
Db      73363 GTCCGCTGCTCGACACAGTGGGCACGGCGGGTACCGGTGATCTGCGCCTGGTGTTCCTCA 73422

Qy       2375 GCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGT 2434
Db      73423 GCGGCGACTTCACGCCGCTGCCCCTCCCCGACGAGGTGCGCGCCGTCTTCCCGCGCGCCG 73482

Qy       2435 CGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGA 2494
Db      73483 ACATGATCAGCCTGGGCGGTGCCACCGAGGCGACGGTGTGGTCCAACTGGTTCCGCATCG 73542

Qy       2495 GGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGT 2554
Db      73543 GGGCGATCGACCCGGCCTGGCGCAGCATCCCCTACGGCAGGCCCATCGACAACAGCCGCT 73602

Qy       2555 TCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACA 2614
Db      73603 ACCACGTGCTCGACGAGGCGCTGCGCCCCTGTCCCGTCGGCGTCGAGGGGGATCTGTACA 73662

Qy       2615 TTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCT 2674
Db      73663 TCGGCGGTGAGTGTCTGGCCCTCGGCTACGTCAACCAGCCCGAGCTGACCGCCGACCGGT 73722

Qy       2675 TCCTC---------GTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGG 2725
Db      73723 TCATCCCCGACCCGTTCCACGAGGACCCTCAGGAGCGCCTCTACAAGACCGGTGACAGGG 73782

Qy       2726 GCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGC 2785
Db      73783 CCCTGTACTACCCGGACGGCAACCTGAGCTTCCAGGGGCGGGCCGACGGACAGGTGAAGG 73842

Qy       2786 TTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACG 2845
Db      73843 TCCGCGGGTTCCGGGTGGAACTGGCCGAGATCGAGCACCGGCTGCGGGCGCACGACGGGG 73902

Qy       2846 TACGCGACGCGGTGATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCT 2905
Db      73903 TCAAGGACGCCGTGGTCCTGGCCCGGGAGGACGGCTGCGGGGACCGCACCCTCGTCGCGT 73962

Qy       2906 ATGTGGT 2912
Db      73963 ACCTCGT 73969


RESULT 9
AAI99682/c
ID   AAI99682 standard; DNA; 4411529 BP.
XX
AC   AAI99682;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
```

```
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

  Query Match             8.5%;  Score 358.2;  DB 22;  Length 4411529;
  Best Local Similarity   48.4%;  Pred. No. 6.6e-66;
  Matches 1252;  Conservative    0;  Mismatches 1283;  Indels   54;  Gaps    7;

Qy        286 CACGCCTATCGCGAATACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTT 345
Db    2675463 CACCTCTACGTCGAGTTCGATGGTGCCCGGGTCGATCCGGACAGGCTCCGCGCGGCGGCC 2675404

Qy        346 CGGAAAGTCGTCGCGCGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAG 405
Db    2675403 ACCCGGTTGGCGCTGCGGCACCCGATGCTGCGGGTGCAGTTCTTGCCCGACGGCACCCAG 2675344

Qy        406 GTGATCGAGCCTAAAGTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGG 465
Db    2675343 CGCATCCCGCCGGCGGCGGGATCTCGCGACTTTCCCATCAGTGTTGCCGACCTGCGTCAC 2675284

Qy        466 AGCACACGGGAAGCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGAC 525
Db    2675283 GTGGCCCCGGATGTCGTCGATCAGCGGTTGGCGGGGATCCGCGACGCCAAATCGCACCAG 2675224

Qy        526 ACCGAGCGCCCTCCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGT 585
Db    2675223 CAGCTCGACGGTGCGGTATTCGAACTTGCGTTGACGTTATTACCGGGGAGCGCACCCGC 2675164

Qy        586 CTCGTGCTCAGTATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAG 645
Db    2675163 CTACATGTCGACCTGGACATGCAGGCCGCTGACGCGATGAGCTACCGCATCTTGCTGGCC 2675104

Qy        646 GACTGGCTCAGCTTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTAC 705
Db    2675103 GACCTGGCGGCCCTCTATGACGGCCGTGAACCGCCGGCACTGGGCTACACCTACCGGGAA 2675044

Qy        706 CGCGATTATGTACTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATG 765
Db    2675043 TACCGGCAGGCTATCGAGGCGGAGGAGACGCTGCCCCAACCGGTTCGCGACGCCGACCGG 2674984

Qy        766 GATTACTGGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCC 825
Db    2674983 GACTGGTGGGCGCAGCGCATCCCGCAGCTGCCGGATCCGCCCGCGCTGCCTACCCGGGCC 2674924

Qy        826 GATCCATCTACCCTGAAGGAGATCCGCTTCGGCACACGGAGCAATGGCTGCCGTCGGAC 885
Db    2674923 GGCGGCGAACGCGACCGGCGCCGCAGCACCCGGCGC---TGGCACTGGCTAGACCCGCAG 2674867

Qy        886 TCCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATC 945
Db    2674866 ACCCGCGACGCGTTGTTCGCCCGAGCCCGGGCCCGCGGCATCACCCCGGCGATGACGCTG 2674807

Qy        946 CTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAAC 1005
Db    2674806 GCCGCGGCCTTTGCCAACGTGCTGGCGCGTTGGTCAGCGTCGTCGCGGTTCCTGCTGAAC 2674747

Qy       1006 ATAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTC 1065
Db    2674746 CTGCCGTTGTTCAGTCGCCAGGCCCTGCATCCAGACGTCGACCTGCTGGTCGGTGACTTC 2674687

Qy       1066 ACGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCT 1125
Db    2674686 ACCTCCTCGCTGTTGCTCGACGTCGATCTGACCGGTGCGCGCACGGCGGCGGCGCGGGCG 2674627

Qy       1126 AAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAG 1185
Db    2674626 CAGGCGGTGCAGGAAGCCCTGCGCAGCGCCGCAGGCCACAGTGCATACCCCGGGCTGTCT 2674567

Qy       1186 GTCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTG 1245
Db    2674566 GTGCTGCGTGACCTCAGCCGCCACCGTGGCACCCAGGTGCTGGCACCGGTGGTATTCACC 2674507

Qy       1246 CTCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACT 1305
Db    2674506 AGCGCGCTGGGGCTCGGCGACCTTTTCTGCCCGGACGTCACCGAGCA---ATTCGGCACA 2674450

Qy       1306 CCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCAC 1365
Db    2674449 CCCGGATGGATCATTTCGCAGGGGCCCCAGGTGCTGCTCGACGCCCAGGTCACCGAGTTC 2674390

Qy       1366 GATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTG 1425
Db    2674389 GACGGCGGTGTGCTGGTGAACTGGGATGTCCGCGAGGGGGTCTTTGCACCCGGCGTCATC 2674330

Qy       1426 GACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCT------CACTGAGGAACCA 1479
Db    2674329 GACGCCATGTTCACCCACCAGGTCGACGAATTGCTCCGGTTGGCCGCCGGGGACGACGCC 2674270

Qy       1480 TGGGGTGAACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCA 1539
Db    2674269 TGGGATGCGCCGAGCCCGTCCGCGCTACCCGCCGCGCAACGCGCGGTGCGCGCGGCGCTG 2674210

Qy       1540 AACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTC 1599
Db    2674209 AACGGTCGCACCGCCGCCCCCAGCACCGAGGCGCTGCACGACGGGTTTTTCCGCCAGGCC 2674150

Qy       1600 GAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAG 1659
Db    2674149 CAACAGCAGCCCGACGCGCCGGCGGTGTTCGCCAGTTCCGGCGACTTGAGCTACGCCCAA 2674090

Qy       1660 CTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAAC 1719
Db    2674089 CTGCGCGACCAGGCATCGGCGGTGGCCGCGGCGCTGCGTGCTGCGGGCCTACGAGTCGGC 2674030

Qy       1720 ACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTG 1779
Db    2674029 GACACCGTCGCGGTGCTGGGTCCGAAAACGGGCGAACAAGTGGCGGCTGTGCTGGGGATT 2673970

Qy       1780 CTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCAC 1839
Db    2673969 TTGGCCGCCGGCGGGGTCTATCTGCCGATCGGCGTCGACCAGCCCCGCGACCGCGCGGAG 2673910

Qy       1840 TACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAA 1899
Db    2673909 CGCATCCTGGCGACCGGTTCGGTCAACTTAGCGCT--------------CGTTTGCGGC 2673865

Qy       1900 CTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGAC 1959
Db    2673864 CCGCCATGCCAAGTGCGGGTGCCGGTCCCGACGCTGTTGCTGGCTGACGTGCTTGCCGCC 2673805

Qy       1960 GGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTAC 2019
Db    2673804 GCGCCGGCAGAATTCGTGCCCGGGCCTAGCGATCCCACCGCGCTCGCCTATGTGTTGTTC 2673745

Qy       2020 ACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAAC 2079
Db    2673744 ACCTCGGGCTCAACCGGGGAACCCAAGGGTGTCGAGGTGGCGCACGACGCCGCGATGAAC 2673685

Qy       2080 ACCATCCTGGACATCAACGAGCGCTTCGAAATAGGCCCGGAGACAGGGTGCTGGCGCTC 2139
Db    2673684 ACCGTGGAGACCTTCATCCGGCACTTCGAGCTAGGCGCCGCAGATCGCTGGCTTGCCCTG 2673625

Qy       2140 TCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGT 2199
Db    2673624 GCGACGCTGGAGTGCGACATGTCGGTGCTGGACATCTTCGCCGCCCTGCGCTCCGGCGGA 2673565

Qy       2200 ACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATC 2259
```

```
Db    2673564 GCGATCGTGGTGGTCGACGAAGCGCAGCGCCGCGATCCCGACGCCTGGGCCCGGCTTATC 2673505

Qy       2260 GAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAG 2319

Db    2673504 GACACTTACGAGGTCACGGCGTTGAATTTCATGCCGGGCTGGTTGGACATGCTGCTCGAA 2673445

Qy       2320 CATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGC 2379

Db    2673444 GTCGGCGGGGGCCGGCTGTCGTCG------------CTGCGAGCGGTGGCCGTCGGCGGC 2673397

Qy       2380 GACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTG 2439

Db    2673396 GACTGGGTGCGTCCCGACCTGGCCCGCCGCCTGCAGGTGCAAGCGCCGAGCGCACGGTTC 2673337

Qy       2440 ATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAAC 2499

Db    2673336 GCGGGGTTGGGTGGAGCCACCGAAACCGCGGTCCACGCAACCATTTTCGAGGTCCAGGAC 2673277

Qy       2500 GTCGACCTATCG---------TGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAG 2550

Db    2673276 GCGGCCAATCTGCCACCGGACTGGGCCTCGGTGCCATACGGCGTCCCGTTTCCCAACAAC 2673217

Qy       2551 ACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTC 2610

Db    2673216 GCCTGCCGGGTAGTGGCCGACAGCGGCGACGACTGCCCCGATTGGGTGGCCGGTGAGTTG 2673157

Qy       2611 TACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAG 2670

Db    2673156 TGGGTGTCCGGTCGCGGAATCGCCCGGGGTTACCGTGGCCGTCCCGAGCTGACCGCGGAG 2673097

Qy       2671 AGCTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGC 2730

Db    2673096 CGCTTCGTCGAGC------ATGACGGCCGCACCTGGTATCGCACCGGTGATCTGGCCCGC 2673043

Qy       2731 TACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGC 2790

Db    2673042 TACTGGCACGACGGCACCCTGGAGTTCGTCGGCCGTGCCGATCACCGCGTCAAGATCAGC 2672983

Qy       2791 GGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGC 2850

Db    2672982 GGGTACCGCGTCGAACTCGGCGAGATCGAAGCCGCGCTGCAGCGCTTGCCCGGTGTGCAC 2672923

Qy       2851 GACGCGGTG 2859

Db    2672922 GCGGCGGCG 2672914


RESULT 10
AAI99683/c
ID   AAI99683 standard; DNA; 4403765 BP.
XX
AC   AAI99683;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

  Query Match             8.2%;  Score 349.2;  DB 22;  Length 4403765;
  Best Local Similarity   50.9%;  Pred. No. 5.5e-64;
  Matches 1073;  Conservative    0;  Mismatches  983;  Indels   54;  Gaps    8;

Qy        765 GGATTACTGGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGC 824

Db    2672280 GGACTGGTGGGCGCAGCGCATCCCGCAGCTGCCGGATCCGCCCGCGCTGCCTACCCGGGC 2672221

Qy        825 CGATCCATCTACCCTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGA 884

Db    2672220 CGGCGGCGAACGCGACCGGCGCCGCAGCACCCGGCGC---TGGCACTGGCTAGACCCGCA 2672164

Qy        885 CTCCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCAT 944

Db    2672163 GACCCGCGACGCGTTGTTCGCCCGAGCCCGGGCCCGCGGCATCACCCCGGCGATGACGCT 2672104

Qy        945 CCTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAA 1004

Db    2672103 GGCCGCGGCCTTTGCCAACGTGCTGGCGCGTTGGTCAGCGTCGTCGCGGTTCCTGCTGAA 2672044

Qy       1005 CATAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTT 1064

Db    2672043 CCTGCCGTTGTTCAGTCGCCAGGCCCTGCATCCAGACGTCGACCTGCTGGTCGGTGACTT 2671984

Qy       1065 CACGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGC 1124

Db    2671983 CACCTCCTCGCTGTTGCTCGACGTCGATCTGACCGGTGCGCGCACGGCGGCGGCGCGGGC 2671924

Qy       1125 TAAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGA 1184

Db    2671923 GCAGGCGGTGCAGGAAGCCCTG-CGCACGCCGCAGGCCACAGTGCATACCCCGGGCTGTC 2671865

Qy       1185 GGTCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGT 1244

Db    2671864 TGTGCTGCGTGACCTCAGCCGCCACCGTGGCACCCAGGTGCTGGCACCGGTGGTATTCAC 2671805

Qy       1245 GCTCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAAC 1304

Db    2671804 CAGCGCGCTGGGGCTCGGCGACCTTTTC--TGCCCGGACGTTCACCGAGCAATTCGGCAC 2671747

Qy       1305 TCCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCA 1364

Db    2671746 ACCCGGATGGATCATTTCGCAGGGGCCCCAGGTGCTGCTCGACGCCCAGGTCACCGAGTT 2671687

Qy       1365 CGATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCT 1424
```

```
Db    2671686 CGACGGCGGTGTGCTGGTGAACTGGGATGTCCGCGAGGGGGTCTTTGCACCCGGCGTCAT 2671627

Qy       1425 GGACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACT------GAGGAACC 1478
Db    2671626 CGACGCCATGTTCACCCACCAGGTCGACGAATTGCTCCGGTTGGCCGCCGGGGACGACGC 2671567

Qy       1479 ATGGGGTGAACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGC 1538
Db    2671566 CTGGGATGCGCCGAGCCCGTCCGCGCTACCCGCCGCGCAACGCGCGGTGCGCGCGGCGCT 2671507

Qy       1539 AAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGT 1598
Db    2671506 GAACGGTCGCACCGCCGCCCCCAGCACCGAGGCGCTGCACGACGGGTTTTTCCGCCAGGC 2671447

Qy       1599 CGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGA 1658
Db    2671446 CCAACAGCAGCCCGACGCGCCGGCGGTGTTCGCCAGTTCCGGCGACTTGAGCTACGCCCA 2671387

Qy       1659 GCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAA 1718
Db    2671386 ACTGCGCGACCAGGCATCGGCGGTGGCCGCGGCGCTGCGTGCTGCGGGCCTACGAGTCGG 2671327

Qy       1719 CACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGT 1778
Db    2671326 CGACACCGTCGCGGTGCTGGGTCCGAAAACGGGCGAACAAGTGGCGGCTGTGCTGGGGAT 2671267

Qy       1779 GCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCA 1838
Db    2671266 TTTGGCCGCCGGCGGGGTCTATCTGCCGATCGGCGTCGACCAGCCCCGCGACCGCGCGGA 2671207

Qy       1839 CTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAA 1898
Db    2671206 GCGCATCCTGGCGACCGGTTCGGTCAACTT--------------AGCGCTCGTTTGCGG 2671162

Qy       1899 ACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGA 1958
Db    2671161 CCCGCCATGCCAAGTGCGGGTGCCGGTCCCGACGCTGTTGCTGGCTGACGTGCTTGCCGC 2671102

Qy       1959 CGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTA 2018
Db    2671101 CGCGCCGGCAGAATTCGTGCCCGGGCCTAGCGATCCCACCGCGCTCGCCTATGTGTTGTT 2671042

Qy       2019 CACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAA 2078
Db    2671041 CACCTCGGGCTCAACCGGGGAACCCAAGGGTGTCGAGGTGGCGCACGACGCCGCGATGAA 2670982

Qy       2079 CACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCT 2138
Db    2670981 CACCGTGGAGACCTTCATCCGGCACTTCGAGCTAGGCGCCGCAGATCGCTGGCTTGCCCT 2670922

Qy       2139 CTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGG 2198
Db    2670921 GGCGACGCTGGAGTGCGACATGTCGGTGCTGGACATCTTCGCCGCCCTGCGCTCCGGCGG 2670862

Qy       2199 TACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGAT 2258
Db    2670861 AGCGATCGTGGTGGTCGACGAAGCGCAGCGCCGCGATCCCGACGCCTGGGCCCGGCTTAT 2670802

Qy       2259 CGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGA 2318
Db    2670801 CGACACTTACGAGGTCACGGCGTTGAATTTCATGCCGGGCTGGTTGGACATGCTGCTCGA 2670742

Qy       2319 GCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGG 2378
Db    2670741 AGTCGGCGGGGGCCGGCTG------------TCGTCGCTGCGAGCGGTGGCCGTCGGCGG 2670694

Qy       2379 CGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCGGCGTGTCGGT 2438
Db    2670693 CGACTGGGTGCGTCCCGACCTGGCCCGCCGCCTGCAGGTGCAAGCGCCGAGCGCACGGTT 2670634

Qy       2439 GATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAA 2498
Db    2670633 CGCGGGGTTGGGTGGAGCCACCGAAACCGCGGTCCACGCAACCATTTTCGAGGTCCAGGA 2670574

Qy       2499 CGTCGAC---------CTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCA 2549
Db    2670573 CGCGGCCAATCTGCCACCGGACTGGGCCTCGGTGCCATACGGCGTCCCGTTTCCCAACAA 2670514

Qy       2550 GACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACT 2609
Db    2670513 CGCCTGCCGGGTAGTGGCCGACAGCGGCGACGACTGCCCCGATTGGGTGGCCGGTGAGTT 2670454

Qy       2610 CTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAA 2669
Db    2670453 GTGGGTGTCCGGTCGCGGAATCGCCCGGGGTTACCGTGGCCGTCCCGAGCTGACCGCGGA 2670394

Qy       2670 GAGCTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCG 2729
Db    2670393 GCGCTTCGTCGAGCATGACGGCCGCAC------CTGGTATCGCACCGGTGATCTGGCCCG 2670340

Qy       2730 CTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCG 2789
Db    2670339 CTACTGGCACGACGGCACCCTGGAGTTCGTCGGCCGTGCCGATCACCGCGTCAAGATCAG 2670280

Qy       2790 CGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACG 2849
Db    2670279 CGGGTACCGCGTCGAACTCGGCGAGATCGAAGCCGCGCTGCAGCGCTTGCCCGGTGTGCA 2670220

Qy       2850 CGACGCGGTG 2859
Db    2670219 CGCGGCGGCG 2670210
```

RESULT 11
AAS53964
ID AAS53964 standard; DNA; 1416 BP.
XX
AC AAS53964;
XX
DT 13-FEB-2002 (first entry)
XX
DE Klebsiella pneumoniae DNA for cellular proliferation protein #93.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Klebsiella pneumoniae.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU36105.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 7601; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to

```
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence encodes an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1416 BP; 271 A; 425 C; 429 G; 291 T; 0 other;

  Query Match              7.2%;  Score 305.6;  DB 23;  Length 1416;
  Best Local Similarity   54.4%;  Pred. No. 1.7e-55;
  Matches  692;  Conservative    0;  Mismatches  559;  Indels   21;  Gaps    3;

Qy       1602 GCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCT 1661
Db         57 GCAACAGCCGCAGGCGCTGGCGGTAACGGACATGCGTTATCAGTGGAATTATCATGAGCT 116

Qy       1662 TTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACAC 1721
Db        117 GACAGACTATGCCCGCCGTTGCGCGGGCAGGTTAATCGAGTGCGGGGTTCAGCCCGGCGA 176

Qy       1722 ATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCT 1781
Db        177 TAATGTGGCTATCACGATGTCGAAAGGCGCAGGACAACTTGTTGCGGTTCTGGCCGTCCT 236

Qy       1782 CGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTA 1841
Db        237 GCTGGCCGGGGCGGTTTACGTTCCGGTTTCGCTGGATCAGCCTGCCGCACGGCGCGAGAA 296

Qy       1842 CCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACT 1901
Db        297 AATCTACGCTGACGCCAGCGTCCGGCTGGTGCTCA------TTTGTCAGCACGACGCCAG 350

Qy       1902 GTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGG 1961
Db        351 CGCCGGGTCAGACGATATTCCCGTCCTTGCCTGGCAGCAGGCCATTGAGGCGGAGCCGAT 410

Qy       1962 CGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACAC 2021
Db        411 CGCCAACCC---------GGTAGTACGCGCCCCCACGCAACCGGCCTACATTATCTACAC 461

Qy       2022 CTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACAC 2081
Db        462 CTCCGGCTCTACCGGTACGCCGAAAGGGGTAGTCATTTCTCACCGGGGAGCGCTTAACAC 521

Qy       2082 CATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTC 2141
Db        522 CTGTTGCGATATCAATACCCGCTATCAGGTTGGCCCGCATGACAGGGTGCTGGCCCTCTC 581

Qy       2142 CTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTAC 2201
Db        582 CGCCCTACATTTTGATTTATCGGTTTACGACATTTTTGGCGTACTGCGCGCGGGCGGCAC 641

Qy       2202 GATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGCAGAGTTGATCGA 2261
Db        642 GCTGGTGATGGTGATGGAAAATCAACGGCGCGATCCTCACGCATGGTGTGAGCTGATCCA 701

Qy       2262 ACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCA 2321
Db        702 GCGCCATCAGGTCACGCTCTGGAACAGCGTCCCGGCGCTGTTCGATATGCTGCTGACCTG 761

Qy       2322 TTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGA 2381
Db        762 GTGTGAAGGTTTCGCCGACGCCACGCCGGAAAACCTGCGCGCAGTGATGCTTTCCGGCGA 821

Qy       2382 CTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGAT 2441
Db        822 CTGGATCGGGCTTGACCTCCCCGCCCGTTATCGGGCCTTCCGGCCACAAGGACAATTTAT 881

Qy       2442 CAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGT 2501
Db        882 CGCGATGGGCGGCGCCACCGAGGCGTCTATCTGGTCTAACGCCTGCGAAATTCACGACGT 941

Qy       2502 CGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGT 2561
Db        942 CCCCGCCCACTGGCGCTCCATCCCTTACGGTTTTCCGCTAACCAACCAACGCTACCGGGT 1001

Qy       2562 GCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGG 2621
Db       1002 GGTGGATGAACAGGGCCGGGACTGCCCTGACTGGGTGCCGGGTGAATTATGGATTGGCGG 1061

Qy       2622 GGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGT 2681
Db       1062 CATTGGGGTCGCGGAAGGCTATTTCAACGATCCCCTGCGTAGCGAGCAGCAATTTTTGAC 1121

Qy       2682 GCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGA 2741
Db       1122 GCTCCCGGA------CGAGCGCTGGTATCGCACCGGCGATCTCGGCTGCTACTGGCCAGA 1175

Qy       2742 TGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGT 2801
Db       1176 TGGCACAATCGAGTTCCTCGGTCGTCGCGACAAGCAGGTCAAAGTCGGAGGATATCGCAT 1235

Qy       2802 TGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGAT 2861
Db       1236 CGAGCTGGGCGAAATCGAAAGCGCGCTCAGCCAGCTGGCGGGGGTGAAACAAGCAACCGT 1295

Qy       2862 TGTGCCCGTCGG 2873
Db       1296 TCTGGCGATCGG 1307

RESULT 12
AAF81367
ID   AAF81367 standard; DNA; 7335 BP.
XX
AC   AAF81367;
XX
DT   04-JUN-2001  (first entry)
XX
DE   Quorum sensing controlled gene qsc109 ORF.
XX
KW   Quorum sensing; antibacterial; bacterial signalling;
KW   opportunistic pathogen; immunocompromised; burn; cystic fibrosis;
KW   immunosuppressive therapy; AIDS; ss.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200118248-A2.
XX
PD   15-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-US24141.
XX
PR   03-SEP-1999;   99US-0153022.
XX
PA   (IOWA ) UNIV IOWA RES FOUND.
PA   (QUOR-) QUORUM SCI INC.
XX
PI   Whiteley M,  Lee KM,  Greenberg EP,  Muh U;
XX
DR   WPI; 2001-265973/27.
XX
PT   Identifying modulators of quorum sensing signaling in Pseudomonas
```

```
PT   aeruginosa bacteria, useful for treating infections in
PT   immunocompromized patients -
XX
PS   Claim 46; Page 108-110; 115pp; English.
XX
CC   Bacteria signal to one another to coordinate expression of specific genes
CC   in a cell density dependent fashion. This "bacterial signalling" is
CC   called "quorum sensing and response". Quorum sensing allows a bacterial
CC   species to sense its own number and regulate gene expression according to
CC   population density. The present sequence is an open reading frame (ORF)
CC   of a Pseudomonas aeruginosa quorum sensing controlled gene. Inhibitors of
CC   quorum sensing signalling renders a bacterial population more susceptible
CC   to treatment. The present invention relates to a method for identifying
CC   modulators of quorum sensing signalling in Pseudomonas aeruginosa
CC   bacteria. Modulators of quorum signaling may be used to treat P.
CC   aeruginosa infections. P. aeruginosa is an opportunistic pathogen of
CC   immunocompromised individuals (burn patients, cystic fibrosis patients,
CC   patients undergoing immunosuppressive therapy and patients with AIDS).
XX
SQ   Sequence 7335 BP; 1184 A; 2256 C; 2625 G; 1270 T; 0 other;

  Query Match             6.7%;  Score 284.2;  DB 22;  Length 7335;
  Best Local Similarity   53.4%;  Pred. No. 9.1e-51;
  Matches  698;  Conservative    0;  Mismatches  588;  Indels   21;  Gaps    4;

Qy       1574 TGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGG 1633
Db       1004 TGCACCGGTTGTTCGAGGAGCAGGTCGAGCGCACGCCGACGGCGCCGGCGCTGGCCTTCG 1063

Qy       1634 CGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGC 1693
Db       1064 GCGAGGAACGCCTGGACTACGCCGAGCTGAACCGCCGGGCCAACCGCCTGGCGCATGCCC 1123

Qy       1694 TGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGG 1753
Db       1124 TGATCGAGCGCGGGGTCGGTCGGACCGCCTGGTGGGCGTGGCCATGGAGCGTTCCATCG 1183

Qy       1754 AGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATG 1813
Db       1184 AGATGGTCGTGGCCCTGATGGCGATCCTCAAGGCCGGCGGCGCCTACGTGCCGGTGGACC 1243

Qy       1814 CCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGC 1873
Db       1244 CGGAGTACCCCGAGGAGCGCCAGGCCTACATGCTGGAGGACAGCGGCGTGCAGCTGCTGC 1303

Qy       1874 TGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCG 1933
Db       1304 TCAGCCAGTCGCACCTGAAGCTGCCGCTGGCGCAAGGCGTGCAGCGGATCGACCTGGACC 1363

Qy       1934 TGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACAC 1993
Db       1364 AGGCCGATGCCTGGCTGGAAAACCATGCCGAGAACAATCCGGGGAT---CGAGCTGAACG 1420

Qy       1994 CTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGTTGCCCAAGGGGGTGA 2053
Db       1421 GCGAGAATCTTGCCTATGTCATCTACACCTCCGGCTCCACCGGCAAGCCCAAGGGTGCCG 1480

Qy       2054 TGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAG 2113
Db       1481 GCAACCGCCATTCGGCGCTGAGCAACCGCTTGTGCTGGATGCAGCAGGCCTACGGCCTGG 1540

Qy       2114 GGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATG 2173
Db       1541 GCGTCGGCGACACGGTGTTGCAGAAGACCCCGTTCAGCTTCGACGTGTCGGTCTGGGAGT 1600

Qy       2174 TGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCG 2233
Db       1601 TCTTCTGGCCGCTGATGAGTGGGGCACGTTTGGTGGTGGCCGCGCCGGGTGACCATCGCG 1660

Qy       2234 ATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGC 2293
Db       1661 ACCCGGCGAAGCTGGTGGCGCTGATCAACCGCGAAGGGGTCGACACGCTGCACTTCGTGC 1720

Qy       2294 CGGCGCTGATGCGGATGCTCGTCGAGCATTTGAGGGTCGCCCCGATTCGCTCGCTAGGT 2353
Db       1721 CGTCGATGCTGCAGGCCTTCCTGCAGGACGAAGACGTCGTCTCCTGCACCAGC------- 1773

Qy       2354 CTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCC 2413
Db       1774 --CTGAAACGCATCGTTTGCAGCGGCGAGGCGCTGTCGGCGGACGCCCAGCAGCAGGTGT 1831

Qy       2414 AGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCT 2473
Db       1832 TCGCCAAGCTGCCGCAGGCCGGCCTCTATAACCTCTATGGCCCGACCGAGGCGGCCAT-- 1889

Qy       2474 GGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCC 2533
Db       1890 ----CGACGTCACCCACTGGAGCTGCGTGGAGGAGGGCAAGGACGCGGTGCCGATCGGCC 1945

Qy       2534 GTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCT 2593
Db       1946 GGCCGATCGCCAACCTGGGCTGCTACATCCTCGATGGCGACCTGGAGCCGGTGCCGGTGG 2005

Qy       2594 GGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATG 2653
Db       2006 GCGTGCTCGGCGAGCTGTACCTGGCCGGTCGGGGCCTGGCTCGTGGCTACCACCAGCGTC 2065

Qy       2654 AAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCC---CGAGACCGGGGAGCGCCTCTACA 2710
Db       2066 CGGGGCTGACTGCCGAGCGTTTCGTCGCCAGCCCGTTCGTGGCTGGGGAGCGGATGTACC 2125

Qy       2711 AGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGG 2770
Db       2126 GCACCGGCGACCTGGCGCGCTACCGCGCCGATGGGGTGATCGAGTACGCCGGGCGGATCG 2185

Qy       2771 ACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCA 2830
Db       2186 ACCACCAGGTGAAGCTGCGCGGCCTGCGCATCGAGCTGGGCGAGATCGAGGCGCGCCTGC 2245

Qy       2831 AGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAAC 2877
Db       2246 TGGAGCATCCGTGGGTGCGCGAGGCGGCGGTGCTGGCGGTGGACAGC 2292


RESULT 13
AAA11992
ID   AAA11992 standard; DNA; 37856 BP.
XX
AC   AAA11992;
XX
DT   07-AUG-2000  (first entry)
XX
DE   S. cellulosum DNA encoding polyketide and hereropolyketide enzymes.
XX
KW   Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;
KW   epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;
KW   plant-protection; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (3398..6100)
FT                   /*tag=  a
FT                   /product= "ORF1-tRNA synthetase"
FT                   /note= "gtg start codon"
FT   CDS             complement (6374..7111)
FT                   /*tag=  b
FT                   /product= "ORF2-monooxygenase"
FT   CDS             complement (8433..9550)
FT                   /*tag=  c
FT                   /product= "ORF3-aminotransferase"
FT                   /note= "AGT start codon given in the specification"
FT   CDS             9855..11393
FT                   /*tag=  d
FT                   /product= "ORF4- tyrosine/DOPA-Decarboxylase"
FT                   /note= "GTG start codon"
```

```
FT   CDS             12212..13658
FT                   /*tag= e
FT                   /product= "ORF5-3-oxoacyl-ACP-reductase"
FT                   /note= "ACC start codon"
FT   CDS             15374..19984
FT                   /*tag= f
FT                   /product= "ORF6-polyketide synthase"
FT   CDS             20003..27889
FT                   /*tag= g
FT                   /product= "ORF7-peptide synthetase"
FT   CDS             28251..29400
FT                   /*tag= h
FT                   /product= "ORF8-transpeptidase"
FT   CDS             complement (30040..31720)
FT                   /*tag= i
FT                   /product= "ORF9-regulation element"
FT                   /note= "CGC stop codon"
FT   CDS             31982..32932
FT                   /*tag= j
FT                   /product= "ORF10-transcription regulator"
FT   CDS             33128..33613
FT                   /*tag= k
FT                   /product= "ORF11-regulation element"
FT                   /note= "GTG start codon"
FT   CDS             33661..34077
FT                   /*tag= l
FT                   /product= "ORF12-regulation element"
FT   CDS             complement (35255..35616)
FT                   /*tag= m
FT                   /product= "ORF13-transcription regulator"
FT   CDS             complement (35730..36242)
FT                   /*tag= n
FT                   /product= "ORF14-transcription regulator"
FT                   /note= "GTG start codon"
XX
PN   DE19846493-A1.
XX
PD   13-APR-2000.
XX
PF   09-OCT-1998;   98DE-1046493.
XX
PR   09-OCT-1998;   98DE-1046493.
XX
PA   (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX
PI   Beyer S,  Mueller R;
XX
DR   WPI; 2000-294101/26.
XX
PT   DNA sequence coding for products involved in the biosynthesis of
PT   polyketide or heteropolyketide compounds, especially epothilone  -
XX
PS   Claim 3; Page 20-33; 36pp; German.
XX
CC   This invention describes a novel DNA sequence (I) whose expression
CC   products effect or are involved in the enzymatic biosynthesis,
CC   mutasynthesis or partial synthesis of polyketide or heteropolyketide
CC   compounds (II). (I) can be inserted into an expression vector and used
CC   to transform or transfect prokaryotic or eukaryotic cells with the aim
CC   of obtaining strains that produce large amounts of polyketide or
CC   heteropolyketide compounds, especially epothilones, which have cytotoxic
CC   and/or immunosuppressant and antibiotic and antifungal activities and
CC   are useful as plant-protection agents. This sequence represents the DNA
CC   sequence isolated from Sorangium cellulosum which is described in the
CC   method of the invention.
XX
SQ   Sequence 37856 BP; 5655 A; 13666 C; 12913 G; 5622 T; 0 other;

  Query Match             6.3%;  Score 265;  DB 21;  Length 37856;
  Best Local Similarity   46.7%;  Pred. No. 1.6e-46;
  Matches 1219;  Conservative    0;  Mismatches 1325;  Indels   69;  Gaps    9;

Qy        401 TGCAGGTGATCGAGCCTAAAGTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCG 460
Db      20259 TGCAGGTGATCGCCCCCTCGCTGCGCATCGAGGTGCCCGTCGTGGACCTGAGCGAGATCG 20318

Qy        461 ACCGGAGCACACGGGAAGCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCT 520
Db      20319 ACGAGCCGGAGCGAGAGGCGGAGGCCCGGCGGCTCATGGCGGAGGAGAGCCGCCGCCCCT 20378

Qy        521 ATGACACCGAGCGCCCTCCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAA 580
Db      20379 TCGATCTCACGCGAGGGCCGCTGCTCCGAGCCAAGCTGCTCCGGCTCGGCGAGGCCGATC 20438

Qy        581 CCCGTCTCGTGCTCAGTATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCT 640
Db      20439 ACGTGCTGATCTTGACGATGCATCATATCGTCTCCGACGGCTGGTCGATGGACGTGCTGT 20498

Qy        641 TCAAGGA------------CTGGCTCAGCTTCTACGAAGATCCCGAGACCTCTCTCCCTG 688
Db      20499 TCAAGGAGCTTTCCACGCTCTACGCCGCCTTCCACGAGGGCCGCCCGTCGCCGCTCCCGG 20558

Qy        689 TCCTGGAGCTCTCGTACCGCGATTATGTACTCGCGCTGGAGTCTCGCAAGAAGTCTGAGG 748
Db      20559 AGCTGCCGATTCAATACGCCGACTTCGCGGTGTGGCAGCGGGAGCTGCTCCAGGGCGAAG 20618

Qy        749 CGCATCAACGATCGATGGATTACTGGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGA 808
Db      20619 TTCTGGAATCGCACCTCGGGTACTGGAGAGAGCACCTCCGCGGCGCCCCCACGCTGCTGG 20678

Qy        809 CGCTTCCGATGAAGGCCGATCCATCTACCCTGAAGGAGATCCGCTTCCGGCACACGGAGC 868
Db      20679 AGCTTCCGATGGACCGGCCCCGGCCGCCGGCGCAGACGTTCCGGGGCTCCCAGCGCGCGT 20738

Qy        869 AATGGCTGCCGTCGGACTCCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGA 928
Db      20739 TCCGACTCCCACTCTCCCTGCAACAGGCGGTGCAGGCGCTCAGCCGGCAGGAAGGCGCGA 20798

Qy        929 CCCCGACGGGCGTCATCCTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCC 988
Db      20799 CCCCCTTCATGACGCTGCTGACGGCGTTCAGCGTGCTGCTCTCGCGTTATGCGCGGCAGA 20858

Qy        989 CCCGGTTTACGCTCAACATAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACG 1048
Db      20859 GCGATCTGGTGGTTGGCACGCCCATCGCGAATCGCACCCGAGCAGAGC------TGGAGG 20912

Qy       1049 ATATCACCGGGGACTTCACGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGA 1108
Db      20913 GGCTGATCGGCTTCTTCGTCAACATGCTGGCGCTGCGCATCGACCTCGGGGGCGACCCGA 20972

Qy       1109 GCTTCGAACAGCGCGCTAAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCG 1168
Db      20973 GCTTCCGCGAGCTGCTCGGGCGGGTGCGGGAGGTGACGTTGGGCGCCTACGCGCACCAGG 21032

Qy       1169 ACGTAAGCGGTATCGAGGTCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCG 1228
Db      21033 ACCTGCCCTTCGAACGGCTGGTGGAGGAGCTGTCACCAGGGCGGAGCCCCAGCCACAGCC 21092

Qy       1229 CATTGTTCCCCGTGGTGCTCACGAGCGCGCTTAACCAGCAAGTCG------TTGGTGTCA 1282
Db      21093 CCTTGTTCCAGGTGTCCTTCACGTTGCAGAACACCCCGATGGATGCGACGAACAGAGCAG 21152

Qy       1283 CCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGC 1342
Db      21153 ACATTGCATCGGGTGGCGCGCCGCTGGTGGAAATGAAGGCGGCGAAATTCGATCTGATCC 21212

Qy       1343 TGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTCCTCGCGTGGGACATCG---TCG 1399
Db      21213 TGGAGCTCTCGGAATCGCCGCAAGGGTTGCTCGGCACGTTCGAGTACAACACCGACCTGT 21272

Qy       1400 ACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAAGCGTACGTCGTTTTCTCC 1459
Db      21273 TCGACGCCGGCACCATCGAGCGGATGGCCGGCCACCTGGAGGTGCTGCTCTCCAGCGCCG 21332

Qy       1460 GGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGC 1519
```

```
Db    21333 TCGCGGCGCCGGATCGACCCATTGCGGAGCTGCCGCTCATGGGGGCCGAGGAGCGCAGTC 21392

Qy     1520 TAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACG 1579
Db    21393 GGGTATTGGTGGAGTGGAACTCCACTGCCGCGCTGTATCCCGAGGACCATTGCATGCACG 21452

Qy     1580 GCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCA 1639
Db    21453 AGCTGTTCGAGCAGCAAGTGGAGCGGTCGCCCGAGGCGACCGCGGTGCTCCTCCAGCAGC 21512

Qy     1640 AGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCG 1699
Db    21513 AGACGTTGACGTATCGAGAGCTGAACATGCGCGCCAATCAGCTCGCGCATCACCTGCGGA 21572

Qy     1700 AGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGG 1759
Db    21573 GCCTGGGCGTGGGCCCAGAGGTGCGCGTCGGGTTGTATCTCGAACGGTCAATCGAGACGG 21632

Qy     1760 TTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACC 1819
Db    21633 TCGTGGCGATCCTCGGCGTGCTCAAGGCTGGCGGGGCCTACGTGCCGCTCGATCCGACGT 21692

Qy     1820 TACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGC 1879
Db    21693 ACCCCAGCGAGCGCCTCGGGCTCATGATGGCGGACGCAGCGCCCTCGGTGCTGCTCACGC 21752

Qy     1880 AGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCG 1939
Db    21753 AGGCGTCGCTCCTCTCGAAGCTGCCGCCCCACGGGGATGCAACGCTGGTACAGCTCGACG 21812

Qy     1940 AGGCCGGCGTCGAAGGCGACGGCGAC---CAGCCTCCGATGATGCCCATTCAGACACCTT 1996
Db    21813 CGCTGCACGAAGCGCTCTCCAGGCTGCCACACCATACCCCGCGGAGCGGCGTCACCGCCC 21872

Qy     1997 CGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGA 2056
Db    21873 AGAACCTCGCATACGTCATGTACACTTCCGGCTCGACCGGGCGGCCCAAGGGCGTGCTCG 21932

Qy     2057 TCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGC 2116
Db    21933 TCGAGCACCGCGGCCTCTGCAACCTGCCCACCGTGCAGGCCAAGCTCTATGGAATCGCGC 21992

Qy     2117 CCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGT 2176
Db    21993 CGGGCGACAGGCTCCTCCAGTTCGCGCCGCTCTGCTTCGACACATCGTTCTGCGAGATCG 22052

Qy     2177 TCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATC 2236
Db    22053 CGCTCGCGTTGCTCTCGGGAGCGACGCTGGTCATGGGCACGGCGGACGAGCTTCTCCCGG 22112

Qy     2237 CGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGG 2296
Db    22113 GACCTCCGCTGGTCGAGCTGCTGAAGAAGCACGCGGTCACGGCGATGCTCCTGGCCCCTA 22172

Qy     2297 CGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTC 2356
Db    22173 CCGTGCTCGCAGCGCTGCCAGAACAACAGAGCGC---------------GGCGTTGCCGC 22217

Qy     2357 TGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGCCTGCCTGGCGAGCTCCAGG 2416
Db    22218 TGCGCGTGCTCACGATGGCCGGTGAGGCGTGCCCGGCGGAGCT------CGTCAAGCGCT 22271

Qy     2417 CCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGT 2476
Db    22272 GGAAGGCACCCGGACGGCGCCTGTTCAACTCCTATGGCCCGACCGAGACGACCATTTGGG 22331

Qy     2477 CCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTC 2536
Db    22332 CAAGCTCCGCAGCGGACCTGTCCGACGAAC---------GGATCCCGCCCATCGGCCGTC 22382

Qy     2537 CGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGG 2596
Db    22383 CGATTGCCAATACGCAAATCTACGTGCTCGACGAAGCGCTCGAGCCGGTGCCCATCGGCG 22442

Qy     2597 TTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAG 2656
Db    22443 TGCCGGGCGAGATCTTCATCGGCGGCGTGGGCGTCGCCCGGGGATATCACGGGCGTCCGG 22502

Qy     2657 AGAAGACGCGCAAGAGCTTCCTCGTGCACCCC---------GAGACCGGGGAGCGCCTCT 2707
Db    22503 ACCTGACGGCCGAGCGATTCGTACCCGACCCCTTCGGGCAAACCAAAGGGGCGCGCCTGT 22562

Qy     2708 ACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTG 2767
Db    22563 ATCGGACCGGCGATCGGGCGCGCTGGCTGCCGGACGGAAACCTCGAGTTTCTCGGTCGAA 22622

Qy     2768 AGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGC 2827
Db    22623 ACGACGAGCAGGTGAAGGTCCGCGGTGTCCGCATCGAGCTGGAGGAGATCCGCGCGGCGT 22682

Qy     2828 TCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGACGCGGCGA 2887
Db    22683 TGCTCAAGCACCCGGCGGTCGCTCAAGCCGTGGCCGTGGTGCGCGAGGACACGCCGGGGG 22742

Qy     2888 ACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCCGAGCAGG 2947
Db    22743 ACAAGCGGCTCGTCGCGTATGTCGTCGGACGCGGAGGAGCGCGCGTGACCGCCGCGGAGC 22802

Qy     2948 ACGCGAGCCTCAAGACCGAGCGGATCGACGCGA 2980
Db    22803 TGCGCCAGTCCGTGAGCGAGCGATTGCCTGCGA 22835
```

RESULT 14
AAS51470

```
ID   AAS51470 standard; DNA; 13029 BP.
XX
AC   AAS51470;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Pseudomonas aeruginosa DNA for cellular proliferation protein #55.
XX
KW   Antisense; ds; prokaryotic cellular proliferation gene;
KW   antibiotic; antibacterial; drug design.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   P-PSDB; AAU33611.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 27; Seq ID No 4052; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
```

```
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence encodes an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 13029 BP; 1978 A; 4677 C; 4485 G; 1889 T; 0 other;

  Query Match             6.2%;  Score 261.4;  DB 23;  Length 13029;
  Best Local Similarity   51.8%;  Pred. No. 7.5e-46;
  Matches  704;  Conservative    0;  Mismatches  631;  Indels   24;  Gaps    4;

Qy       1572 GCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTC 1631
Db       3387 GCTGCCGGAGCTGCTGGAGCGGCAACTGGCGCAGTCCGCCGAACGCGTCGCCCTGGAGTG 3446

Qy       1632 GGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCG 1691
Db       3447 GGACGGCGGCAGCCTCGGCTACGCCGAACTGCACGCCCGGGCCAACCGCCTGGCCCACTA 3506

Qy       1692 GCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTG 1751
Db       3507 CCTGCGCGACAAGGGCGTCGGCCCGGACGTGCGGGTAGCGATCTGCGCCGAGCGTTCGCC 3566

Qy       1752 GGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGA 1811
Db       3567 GCAACTGCTGGTCGGTCTGCTGGCGATCGTCAAGGCGGGCGGTGCCTACGTGCCGCTGGA 3626

Qy       1812 TGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGT 1871
Db       3627 CCCGGACTACCCCAGCGAACGCCTGGCCTACATGCTCGCCGACAGTGGCGTCGAGCTGCT 3686

Qy       1872 GCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCT 1931
Db       3687 ACTGACCCAGGCCCATCTGTTCGAACGCCTGCCTGGCGCGGAGGGCGTCACGCCGATCTG 3746

Qy       1932 CGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGAC 1991
Db       3747 CCTGGACAGCCTCAAGCTCGACAACTGGCCGAGCCAGGCACCGGGCCTGCACCTGCACGG 3806

Qy       1992 ACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGT 2051
Db       3807 CGACA---ATCTCGCCTACGTCATCTACACCTCCGGCTCCACCGGCCAGCCGAAGGGCGT 3863

Qy       2052 GATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAAT 2111
Db       3864 GGGCAACACCCACGCGGCGCTGGCCGAGCGCCTGCAATGGATGCAGGCCACCTACACCCT 3923

Qy       2112 AGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGA 2171
Db       3924 CGACGGCGACGACGTGCTGATGCAGAAGGCGCCGGTGAGCTTCGACGTGTCGGTGTGGGA 3983

Qy       2172 TGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCG 2231
Db       3984 GTGCTTCTGGCCGCTGGTCACCGGTTGCCGGTTGGTCCTCGCCGCCCCCGGCGAGCACCG 4043

Qy       2232 CGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGT 2291
Db       4044 CGATCCGGCGCGCCTGGTGGAACTGGTGCGGCAGTTCGGGGTGACCACGCTGCACTTCGT 4103

Qy       2292 GCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAG 2351
Db       4104 GCCGCCGCTGCTGCAACTGTTCATC---------GACGAACCCGGCGTCGCCGCCTGCGG 4154

Qy       2352 GTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCT 2411
Db       4155 CAGCCTGCGCCGCCTGTTCTCCGGCGGCGAGGCGCTTCCGGCGGAGCTGCGCAACCGCGT 4214

Qy       2412 CCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGAT 2471
Db       4215 GTTGCAACGCCTGCCGGCGGTGGCCCTGCATAACCGCTACGGGCCGACCGAGACCGCCAT 4274

Qy       2472 CTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGG 2531
Db       4275 ------CAACGTCACCCATTGGCAGTGCCGCGCGGAAGACGGCGAGCGTTCGCCGATCGG 4328

Qy       2532 CCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGT 2591
Db       4329 CCGACCGCTGGGCAACGTGGTCTGCCGCGTGCTGGACGCCGAGTTCAACCTGCTGCCGGC 4388

Qy       2592 CTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGA 2651
Db       4389 CGGCGTCGCCGGCGAGCTGTGCATCGGCGGCCTGGGCCTGGCCCGCGGCTACCTGGGGCG 4448

Qy       2652 TGAAGAGAAGACGCGCAAGAGCTTCCTCG------TGCACCCCGAGACCGGGGAGCGCCT 2705
Db       4449 CCCGGCGCTGAGCGCCGAGCGCTTCGTCGCCGATCCGTTCTCGGCCGCCGGCGAGCGCCT 4508

Qy       2706 CTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCG 2765
Db       4509 GTACCGCACCGGCGACCGCGCGCGCTGGAACGCCGACGGCGTGCTGGAATACCTCGGCCG 4568

Qy       2766 TGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAAC 2825
Db       4569 CCTCGACCAGCAGGTCAAGCTGCGTGGCTTCCGCATCGAGCCGGAGGAAATCCAGGCACG 4628

Qy       2826 GCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGACGCGGC 2885
Db       4629 CCTGCTGGCGCAGCCGGGGGTGGCCCAGGCGGTAGTGGTGATCCGCGAAGGCGTGGCCGG 4688

Qy       2886 GAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCAC 2924
Db       4689 CAGCCAGTTGGTCGGCTACTACACCGGCGCTGTCGGCGC 4727

RESULT 15
ABQ78872
ID   ABQ78872 standard; DNA; 90600 BP.
XX
AC   ABQ78872;
XX
DT   23-OCT-2002  (first entry)
XX
DE   S. roseosporus daptomycin biosynthetic gene cluster 90kb region.
XX
KW   Daptomycin biosynthetic gene cluster; thioesterase; antibacterial;
KW   fungicide; virucide; antiparasitic; immunomodulator; antilipemic;
KW   cytostatic; gene therapy; antimitotic; immunomodulatory; siderophore;
KW   anti-cholesterolemic; agrochemical; gene; ds.
XX
OS   Streptomyces roseosporus.
XX
PN   WO200259322-A2.
XX
PD   01-AUG-2002.
XX
PF   17-OCT-2001; 2001WO-US32354.
XX
PR   17-OCT-2000; 2000US-240879P.
PR   28-FEB-2001; 2001US-272207P.
PR   06-AUG-2001; 2001US-310385P.
XX
PA   (MIAO/) MIAO V P W.
```

```
PA   (BRIA/) BRIAN P.
PA   (BALT/) BALTZ R H.
PA   (SILV/) SILVA C J.
XX
PI   Miao VPW,  Brian P,  Baltz RH,  Silva CJ;
XX
DR   WPI; 2002-599794/64.
XX
PT   Isolated nucleic acid molecule from a bacterial daptomycin biosynthetic
PT   gene cluster encoding a thioesterase or thioesterase domain, useful for
PT   generating novel linear and cyclic peptides, and products in a cell  -
XX
PS   Claim 7; Page 142-165; 227pp; English.
XX
CC   The invention relates to a novel isolated nucleic acid molecule
CC   comprising a sequence that encodes a thioesterase or thioesterase domain,
CC   derived from a bacterial daptomycin biosynthetic gene cluster. The
CC   proteins of the invention have antibacterial, fungicide, virucide,
CC   antiparasitic, immunomodulator, antilipemic, and cytostatic activity. The
CC   polynucleotides may have a use in gene therapy. The compositions and
CC   methods of the present invention are useful for generating novel linear
CC   and cyclic peptides and improving yield of a product in a cell expressing
CC   an daptomycin non-ribosomal peptide synthetase (NRPS) to be used as new
CC   compounds or in producing new compounds, such as antibiotics,
CC   antifungals, antivirals, antiparasitics, antimitotics, antitumour agents,
CC   immunomodulatory agents, anti-cholesterolemic agents, siderophores,
CC   agrochemicals and cytostatics. The sequence represents the 90kb region of
CC   the S. roseosporus daptomycin biosynthetic gene cluster.
XX
SQ   Sequence 90600 BP; 12671 A; 32312 C; 31571 G; 14046 T; 0 other;

  Query Match             5.8%;  Score 247.6;  DB 24;  Length 90600;
  Best Local Similarity   51.4%;  Pred. No. 1e-42;
  Matches  711;  Conservative    0;  Mismatches  649;  Indels   24;  Gaps    5;

Qy       1535 GCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGC 1594
Db      54165 GGGGCAGCGGTACGCACACCCCGCTGCCCCCGCGCAGCGTGGCCGAGCAGATCGTCCGCC 54224

Qy       1595 GGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACG 1654
Db      54225 GGGCCGCGCTGGACCCGGACGCCGTCGCCGTCATCACCGCGGAAGAGGAACTCTCGTACC 54284

Qy       1655 AAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCC 1714
Db      54285 GGGAACTGGAAAGGCTCAGCGGTGAGACGGCGCGGCTGCTGGCCGACCGGGGGATCGGCC 54344

Qy       1715 CGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCG 1774
Db      54345 GCGAGAGCCTCGTCGCCGTCGCCCTGCCCCGCACGGCCGGCCTGGTCACCACCCTGCTCG 54404

Qy       1775 CGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTA 1834
Db      54405 GCGTCCTGCGCACCGGCGCCGCCTACCTCCCGCTCGACACCGGGTACCCCGCCGAGCGAC 54464

Qy       1835 TCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATG 1894
Db      54465 TCGCGCACGTGCTCTCCGACGCCCGTCCCGACCTCGTCCTCACCCACGCCGGCCTCGCCG 54524

Qy       1895 GCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAG 1954
Db      54525 GACGGCT------GCCGGCCGGCCTCGCGCCGACCGTCCTCGTCGACGAGCCGCAGCCGC 54578

Qy       1955 GCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCA 2014
Db      54579 CCGCCGCAGCCGCCCCCGCGGTTCCCACGTCCCCGTCGGGCGACCACCTCGCGTACGTCA 54638

Qy       2015 TCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCG 2074
Db      54639 TCCACACCTCCGGCTCCACCGGCAGGCCCAAGGGCGTCGCGATCGCCGAGTCCTCCCTGC 54698

Qy       2075 TCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGG 2134
Db      54699 GCGCCTTCCTCGCGGACGCGGTCCGGCGCCACGACCTGACCCCGCACGACCGGTTGCTCG 54758

Qy       2135 CGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGG 2194
Db      54759 CGGTGACCACCGTCGGCTTCGACATCGCCGGCCTCGAACTGTTCGCCCCGCTCCTCGCCG 54818

Qy       2195 GCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGT 2254
Db      54819 GTGCCGCGATCGTGCTGGCCGACGAGGACGCCGTACGCGACCCCGCCTCGATCACCTCCC 54878

Qy       2255 TGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCG 2314
Db      54879 TGTGCGCACGCCACCACGTCACCGTCGTCCAGGCCACGCCCAGTTGGTGGCGGGCCATGC 54938

Qy       2315 TCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGA 2374
Db      54939 TCGACGGAGCACCGGCCGACGCCGCCGCCCGGCTCGAGCACGTACGGATC---CTGGTCG 54995

Qy       2375 GCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGT 2434
Db      54996 GCGGCGAACCGCTGCCCGCCGACCTGGCCCGTGTCCTGACCGCAA------CCGGCGCCG 55049

Qy       2435 CGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGA 2494
Db      55050 CCGTCACCAACGTGTACGGACCCACCGAAGCCACCATCTGGGCCACCGCCGCCCCACTCA 55109

Qy       2495 GGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGT 2554
Db      55110 CCGCCGGCGACGACCGCACACCCGGCAT------CGGCACCCCCCTGGACAACTGGCGCG 55163

Qy       2555 TCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACA 2614
Db      55164 TCCACATACTCGACGCGGCCCTCGGACCCGTTCCCCCGGGTGTTCCGGGCGAGATCCACA 55223

Qy       2615 TTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCT 2674
Db      55224 TCGCCGGGTCCGGGCTCGCCCGCGGCTATCTGCGCCGCCCGGACCTCACCGCCGAACGCT 55283

Qy       2675 TCCTCGTGCACCC---CGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCT 2731
Db      55284 TCGTCGCCAACCCGTTCGCCCCCGGCGAGCGGATGTACCGCACCGGCGACCTCGGCCGGT 55343

Qy       2732 ACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCG 2791
Db      55344 TCCGCCCGGACGGCACGCTCGAACACCTCGGCCGCGTGGACGACCAGGTCAAGGTACGGG 55403

Qy       2792 GATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCG 2851
Db      55404 GCTTCCGCATCGAACTCGGCGACGTCGAGGCCGCCCTCGCCCGGCATCCCGACGTGGGGC 55463

Qy       2852 ACGCGGTGATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGG 2911
Db      55464 GCGCCGCCGCCGCCGTCCGCCCCGACCACCGCGGCCAGGGCCGCCTTGTCGCGTACGTCG 55523

Qy       2912 TCCC 2915
Db      55524 TCCC 55527

Search completed: October  3, 2003, 12:28:49
Job time : 1085.8 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:          October  3, 2003, 11:40:41 ; Search time 15142.8 Seconds
                                           (without alignments)
                                           11435.853 Million cell updates/sec

Title:           US-10-014-717-1_COPY_11872_16104
Perfect score:   4233
Sequence:        1 atgacgatcaatcagcttct..........agggcaggagacgtagctaa 4233

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters:      5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                 1:  gb_ba:*
                 2:  gb_htg:*
                 3:  gb_in:*
                 4:  gb_om:*
                 5:  gb_ov:*
                 6:  gb_pat:*
                 7:  gb_ph:*
                 8:  gb_pl:*
                 9:  gb_pr:*
                 10:  gb_ro:*
                 11:  gb_sts:*
                 12:  gb_sy:*
                 13:  gb_un:*
                 14:  gb_vi:*
                 15:  em_ba:*
                 16:  em_fun:*
                 17:  em_hum:*
                 18:  em_in:*
                 19:  em_mu:*
                 20:  em_om:*
                 21:  em_or:*
                 22:  em_ov:*
                 23:  em_pat:*
                 24:  em_ph:*
                 25:  em_pl:*
                 26:  em_ro:*
                 27:  em_sts:*
                 28:  em_un:*
                 29:  em_vi:*
                 30:  em_htg_hum:*
                 31:  em_htg_inv:*
                 32:  em_htg_other:*
                 33:  em_htg_mus:*
                 34:  em_htg_pln:*
                 35:  em_htg_rod:*
                 36:  em_htg_mam:*
                 37:  em_htg_vrt:*
                 38:  em_sy:*
                 39:  em_htgo_hum:*
                 40:  em_htgo_mus:*
                 41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4233	100.0	68750	1	AF210843	AF210843 Sorangium
	2	4233	100.0	68750	6	AR193029	AR193029 Sequence
	3	4233	100.0	68750	6	AR199551	AR199551 Sequence
	4	4233	100.0	68750	6	AR199559	AR199559 Sequence
	5	4233	100.0	68750	6	AR199567	AR199567 Sequence
	6	4233	100.0	68750	6	AR201097	AR201097 Sequence
	7	4233	100.0	68750	6	AR208671	AR208671 Sequence
	8	4114.6	97.2	58733	1	AF217189	AF217189 Sorangium
	9	4114.6	97.2	71989	6	AR172664	AR172664 Sequence
	10	944.8	22.3	42603	1	AF188287	AF188287 Stigmatel
	11	728	17.2	77457	1	AF210249	AF210249 Streptomy
c	12	634.6	15.0	189050	1	AL646066	AL646066 Ralstonia
	13	609.4	14.4	40156	1	AF516145	AF516145 Lyngbya m
	14	600	14.2	302007	1	SCO939132	AL939132 Streptomy
	15	593.6	14.0	189050	1	AL646066	AL646066 Ralstonia
c	16	577.6	13.6	18023	6	AX024285	AX024285 Sequence
c	17	550.6	13.0	311143	1	AE016762	AE016762 Escherich
	18	543.4	12.8	4305	6	AX276331	AX276331 Sequence
	19	543.4	12.8	12303	1	AE009297	AE009297 Agrobacte
c	20	543.4	12.8	31220	1	AE008316	AE008316 Agrobacte
c	21	543.4	12.8	48715	6	AX276312	AX276312 Sequence
c	22	543.4	12.8	50538	6	AX702471	AX702471 Sequence
	23	538.8	12.7	5430	1	AF184622	AF184622 Pseudomon
c	24	526	12.4	17071	1	AE004839	AE004839 Pseudomon
	25	508.4	12.0	15023	1	AF074705	AF074705 Pseudomon
	26	476.8	11.3	290850	1	SCO939127	AL939127 Streptomy
	27	454.4	10.7	4317	1	AF184621	AF184621 Pseudomon
c	28	406.6	9.6	309904	1	AE016865	AE016865 Pseudomon
	29	395	9.3	7444	1	AE013843	AE013843 Yersinia
	30	395	9.3	35806	1	AF091251	AF091251 Yersinia
	31	395	9.3	119443	1	YP102KB	AL031866 Yersinia
	32	395	9.3	311143	1	AE016762	AE016762 Escherich
c	33	395	9.3	335050	1	AJ414150	AJ414150 Yersinia
	34	390.4	9.2	32159	6	AX702468	AX702468 Sequence
	35	383.6	9.1	6468	1	YEPIRP2H	L18881 Yersinia en
	36	360.6	8.5	135638	1	AF484556	AF484556 Streptomy
c	37	358.2	8.5	10570	1	MTCY22H8	Z81371 Mycobacteri
c	38	358.2	8.5	19782	1	AE007084	AE007084 Mycobacte
c	39	358.2	8.5	306550	1	BX248342	BX248342 Mycobacte
c	40	333.8	7.9	41156	1	MSGY222	AD000010 Mycobacte
	41	329.2	7.8	276800	1	SCO939115	AL939115 Streptomy
c	42	317.4	7.5	216050	1	AL646076	AL646076 Ralstonia
	43	296.8	7.0	23451	1	AB070952	AB070952 Streptomy
c	44	296.8	7.0	302300	1	AP005034	AP005034 Streptomy
c	45	287.2	6.8	15719	1	AE004667	AE004667 Pseudomon

ALIGNMENTS

```
RESULT 1
AF210843
LOCUS       AF210843                68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
```

```
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /mol_type="genomic DNA"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
                     /db_xref="GI:6724239"
                     /translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
                     AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
                     EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
                     SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
                     VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
                     GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
                     TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
                     IGIVMTQSAGFLFSGALERFWRSVYVATESA"
     CDS             3415. .5556
                     /note="ORF3; similar to Na/H and drug/H antiporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26906.1"
                     /db_xref="GI:6724240"
                     /translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
                     GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
                     RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
                     RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
                     LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
                     NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
                     LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
                     SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
                     IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
                     IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
                     ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
                     ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
                     YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
     CDS             complement(5612. .5992)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26907.1"
                     /db_xref="GI:6724241"
                     /translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTIAYTDRE"
     CDS             6226. .6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MKHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYANFELLITAEKQSSPQSPSSAAVIGPTSVG"
     gene            7610. .11875
                     /gene="epoA"
     CDS             7610. .11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
     misc_feature    7640. .8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
     misc_feature    9236. .10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
     misc_feature    10529. .11428
                     /gene="epoA"
                     /note="Region: enoyl reductase"
     misc_feature    11549. .11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
     gene            11872. .16104
                     /gene="epoP"
     CDS             11872. .16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
```

```
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPVTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
     gene            16251..21749
                     /gene="epoB"
     CDS             16251..21749
                     /gene="epoB"
                     /note="EPOS B"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26920.1"

  Query Match             100.0%;  Score 4233;  DB 1;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy       1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy       1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy       1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy       1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371
```

```
Qy       1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy       1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy       1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy       1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy       1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy       1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy       1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy       1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy       1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy       2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy       2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy       2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy       2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy       2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy       2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy       2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy       2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy       2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy       2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGCTGGCACTGGGC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGCTGGCACTGGGC 14511

Qy       2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy       2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy       2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy       2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy       2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy       2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy       3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy       3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy       3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy       3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy       3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy       3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy       3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy      3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy      3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy      3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy      3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy      3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy      4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy      4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy      4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy      4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
             |||||||||||||||||||||||||||||||||
Db     16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

```
RESULT 2
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 4233;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy        61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy       121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy       181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy       241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy       301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy       361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy       421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy       481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy       541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy       601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy       661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy       721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy       781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy       841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy       901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy       961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy      1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy      1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy      1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy      1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy      1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy      1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy      1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy      1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy      1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy      1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy      1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy      1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy      1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy      1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy      1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy      1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy      1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy      2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy      2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy      2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy      2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy      2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGTCGCCCCGAT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGTCGCCCCGAT 14211

Qy      2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy      2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy      2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy      2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy      2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy      2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy      2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy      2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy      2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy      2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy      2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy      3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy      3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy      3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy      3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy      3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy      3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy      3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291
```

```
Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||||||||||||||||||||||||||||||||
Db      16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

```
RESULT 3
AR199551
LOCUS       AR199551                68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 4233;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831
```

```
Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy       1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy       1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy       1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy       1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy       1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy       1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy       1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy       1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy       1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy       1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy       1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy       1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy       1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy       2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy       2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy       2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy       2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy       2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy       2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy       2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy       2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy       2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy       2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy       2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy       2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy       2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy       2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy       2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy       2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy       3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy       3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy       3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy      3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy      3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy      3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy      3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy      3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy      3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy      3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy      3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy      3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy      3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy      3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy      3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy      3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy      3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy      4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy      4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy      4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy      4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
             |||||||||||||||||||||||||||||||||
Db     16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

```
RESULT 4
AR199559
LOCUS       AR199559                68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 4233;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy        61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy       121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy       181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy       241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy       301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy       361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy       421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy       481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy       541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy       601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy       661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy       1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy       1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy       1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy       1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy       1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy       1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy       1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy       1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy       1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy       1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy       1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy       1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy       1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy       2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy       2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy       2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy       2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy       2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy       2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy       2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy       2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy       2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy       2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy       2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy       2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy       2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy       2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751
```

```
Qy       2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy       2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy       3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy       3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy       3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy       3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy       3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy       3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy       3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||||||||||||||||||||||||||||||||
Db      16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104


RESULT 5
AR199567
LOCUS       AR199567                68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 4233;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12232 CGGCACGAGATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291
```

```
Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy       1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy       1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy       1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy       1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy       1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy       1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy       1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy       1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy       1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy       1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy       1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy       1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy       1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy       2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy       2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy       2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy       2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy       2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy       2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy       2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy       2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy       2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy       2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy      2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy      2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy      2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy      2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy      2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy      2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy      3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy      3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy      3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy      3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy      3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy      3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy      3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy      3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy      3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy      3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy      3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy      3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy      3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy      3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy      3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy      3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy      3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy      4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy      4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy      4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy      4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
             |||||||||||||||||||||||||||||||||
Db     16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104


RESULT 6
AR201097
LOCUS       AR201097                68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 4233;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy      181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy      241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy      301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy      361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy      421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy      481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy      541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy      601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy      661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy      721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy      781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy      841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy      901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy      961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy     1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy     1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy     1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy     1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy     1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy     1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy     1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy     1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy     1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy     1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy     1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy     1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy     1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy     1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy     1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy     1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy     1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy     2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy     2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy     2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy     2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy     2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTGAGGGTCGCCCCGAT 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTGAGGGTCGCCCCGAT 14211
```

```
Qy       2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy       2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy       2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy       2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy       2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy       2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy       2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy       2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy       2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy       2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy       2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy       3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy       3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy       3121 GACGTCTACGCGCGTCGCCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14992 GACGTCTACGCGCGTCGCCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy       3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy       3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy       3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy       3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy       3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
              |||||||||||||||||||||||||||||||||
Db      16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

```
RESULT 7
AR208671
LOCUS       AR208671                68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
```

```
                    /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 4233;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 4233;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11872 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 11931

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11932 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 11991

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11992 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 12051

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12052 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 12111

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12112 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 12171

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12172 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 12231

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12232 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 12291

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 12351

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12352 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 12411

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12412 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 12471

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12472 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12532 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 12591

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12592 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 12651

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12652 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 12711

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12712 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 12771

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12772 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 12831

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12832 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 12891

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12892 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 12951

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12952 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 13011

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13012 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 13071

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13072 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 13131

Qy       1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13132 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 13191

Qy       1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13192 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 13251

Qy       1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13252 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 13311

Qy       1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13312 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 13371

Qy       1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13372 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 13431

Qy       1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13432 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 13491

Qy       1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13492 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 13551

Qy       1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13552 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 13611

Qy       1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13612 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 13671

Qy       1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13672 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 13731

Qy       1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13732 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 13791

Qy       1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13792 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 13851

Qy       1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13852 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 13911

Qy       2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     13912 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 13971

Qy      2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13972 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 14031

Qy      2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14032 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 14091

Qy      2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14092 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 14151

Qy      2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14152 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 14211

Qy      2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14212 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 14271

Qy      2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14272 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 14331

Qy      2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14332 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 14391

Qy      2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14392 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 14451

Qy      2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14452 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 14511

Qy      2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14512 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 14571

Qy      2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14572 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 14631

Qy      2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14632 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 14691

Qy      2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14692 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 14751

Qy      2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14752 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 14811

Qy      2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14812 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 14871

Qy      3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14872 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 14931

Qy      3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14932 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 14991

Qy      3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14992 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 15051

Qy      3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 15111

Qy      3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15112 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 15171

Qy      3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15172 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 15231

Qy      3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15232 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 15291

Qy      3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15292 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 15351

Qy      3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15352 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 15411

Qy      3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15412 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 15471

Qy      3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15472 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 15531

Qy      3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15532 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 15591

Qy      3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15592 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 15651

Qy      3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15652 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 15711

Qy      3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15712 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 15771

Qy      3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15772 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 15831

Qy      3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15832 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 15891

Qy      4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15892 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 15951

Qy      4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     15952 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 16011

Qy      4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     16012 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 16071

Qy      4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
             |||||||||||||||||||||||||||||||||
Db     16072 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 16104
```

```
RESULT 8
AF217189
LOCUS       AF217189               58733 bp    DNA     linear   BCT 09-JUN-2000
DEFINITION  Sorangium cellulosum putative transposase gene, partial cds;
            putative transposase gene, complete cds; epothilone biosynthesis
            gene cluster, complete sequence; putative membrane protein gene,
            complete cds.
ACCESSION   AF217189
VERSION     AF217189.1  GI:7453554
KEYWORDS    .
SOURCE      Polyangium cellulosum
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 58733)
  AUTHORS   Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and
            Julien,B.
  TITLE     Cloning and heterologous expression of the epothilone gene cluster
  JOURNAL   Science 287 (5453), 640-642 (2000)
  MEDLINE   20115953
   PUBMED   10649995
REFERENCE   2  (bases 1 to 58733)
  AUTHORS   Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
  TITLE     Isolation and characterization of the epothilone biosynthetic gene
            cluster from Sorangium cellulosum
  JOURNAL   Gene 249 (1-2), 153-160 (2000)
  MEDLINE   20293058
   PUBMED   10831849
REFERENCE   3  (bases 1 to 58733)
  AUTHORS   Julien,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
     source          1. .58733
                     /organism="Polyangium cellulosum"
                     /mol_type="genomic DNA"
                     /strain="SMP44"
                     /db_xref="taxon:56"
     CDS             <1. .992
                     /note="orfA"
                     /codon_start=3
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62878.1"
                     /db_xref="GI:7453555"
                     /translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
                     SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
                     DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
                     DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
                     GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
                     HIGAVRQVIDRRRSERHLPPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"
     CDS             989. .1501
                     /note="orfB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62879.1"
                     /db_xref="GI:7453556"
                     /translation="MTDLTPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
                     RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
                     AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
                     RTKQRRTKKH"
     gene            1998. .6263
                     /gene="epoA"
     CDS             1998. .6263
                     /gene="epoA"
                     /note="type I polyketide synthase; epothilone PKS loading
                     module"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoA"
                     /protein_id="AAF62880.1"
                     /db_xref="GI:7453557"
                     /translation="MADRPIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
                     GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
                     PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
                     DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
                     QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"
     gene            6260. .10492
                     /gene="epoB"
     CDS             6260. .10492
                     /gene="epoB"
                     /note="non ribosomal peptide synthetase; epothilone PKS
                     module 1"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoB"
                     /protein_id="AAF62881.1"
                     /db_xref="GI:7453558"
                     /translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
                     PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
                     ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
                     STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
                     FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"
     gene            10639. .16137
                     /gene="epoC"
     CDS             10639. .16137
                     /gene="epoC"
                     /note="type I polyketide synthase; epothilone PKS module
                     2"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoC"
                     /protein_id="AAF62882.1"
```

```
                     /db_xref="GI:7453559"
                     /translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
                     ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
                     ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
                     HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
                     AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
                     GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
                     DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
                     ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
                     FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
                     REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
                     SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
                     VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
                     GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
                     EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
                     GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
                     VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene            16134..37907
                     /gene="epoD"
     CDS             16134..37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                     GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                     FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                     RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                     LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                     LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                     DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                     AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                     NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                     AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                     GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                     AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                     RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA

  Query Match             97.2%;  Score 4114.6;  DB 1;  Length 58733;
  Best Local Similarity   98.3%;  Pred. No. 0;
  Matches 4159;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
Db       6260 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTGTCAAGCTGGCGGCCGAT 6319

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
Db       6320 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 6379

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
Db       6380 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAGTCC 6439

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
Db       6440 ATCGTGCCCGCCCCAGCCGAGCGGCACGTTCCGTTTCCTCTCACAGACATCCAAGGATCC 6499

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
Db       6500 TACTGGCTGGGTCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 6559

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
Db       6560 TACGACTGTACGGATCTCGACGTGGCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 6619

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
Db       6620 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 6679

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
Db       6680 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 6739

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
Db       6740 AGGCTCGTATCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 6799

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
Db       6800 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCAGCAAACCCGTCTCGTGCTCAGTATC 6859

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
Db       6860 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGATTGGCTCAGCTTC 6919

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
Db       6920 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGACTATGTGCTC 6979

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
Db       6980 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 7039

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
Db       7040 CGCGTCGCCGAGCTCCCACCTCCGCCGATGCTTCCGATGAAGGCCGATCCATCTACCCTG 7099

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
Db       7100 AGGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGAGTCGATTG 7159

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
Db       7160 AAGCAGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATTCTGGCTGCATTTTCC 7219

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
Db       7220 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 7279

Qy       1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
Db       7280 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 7339

Qy       1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
Db       7340 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 7399

Qy       1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
Db       7400 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 7459

Qy       1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
Db       7460 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTC 7519
```

```
Qy   1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
Db   7520 AACCAGCAAGTCGTTGGTGTCACCTCGCTGCAGAGGCTCGGCACTCCGGTGTACACCAGC 7579

Qy   1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
Db   7580 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 7639

Qy   1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
Db   7640 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 7699

Qy   1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
Db   7700 GCGTACGTCGCTTTTCTCCGGCGGCTCACTGAGGAACCATGGAGTGAACAGATGCGCTGT 7759

Qy   1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
Db   7760 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGAGACCAACTCGCTGCTG 7819

Qy   1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
Db   7820 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCTATGCAGCTC 7879

Qy   1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
Db   7880 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 7939

Qy   1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
Db   7940 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 7999

Qy   1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
Db   8000 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 8059

Qy   1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
Db   8060 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 8119

Qy   1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
Db   8120 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 8179

Qy   1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
Db   8180 CAGCGGCTGCTCGTGAGCGATGCCGGCGTCGAAGGCGACGGCGACCAGCTTCCGATGATG 8239

Qy   1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
Db   8240 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 8299

Qy   2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
Db   8300 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 8359

Qy   2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
Db   8360 CGCTTCGAAATAGGGCCCGGAGACAGAGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 8419

Qy   2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
Db   8420 TCGGTCTACGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 8479

Qy   2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
Db   8480 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGCGTTGATCGAACGAGAGAAGGTGACGGTG 8539

Qy   2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
Db   8540 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTCCGAGGGTCGCCCCGAT 8599

Qy   2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
Db   8600 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 8659

Qy   2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
Db   8660 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 8719

Qy   2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
Db   8720 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGATCCATCGTGGGCGAGC 8779

Qy   2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
Db   8780 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 8839

Qy   2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
Db   8840 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGACTGGCACTGGGC 8899

Qy   2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
Db   8900 TACTGGCGCGATGAAGAGAAGACGCGCAACAGCTTCCTCGTGCACCCCGAGACCGGGGAG 8959

Qy   2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
Db   8960 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 9019

Qy   2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
Db   9020 GGGCGGGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 9079

Qy   2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
Db   9080 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 9139

Qy   2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
Db   9140 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAAGGCACACGGAGACGCGCTGCC 9199

Qy   2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
Db   9200 GAGCAGGACGCGAGCCTCAAGACCGAGCGGGTCGACGCGAGAGCACACGCCGCCAAAGCG 9259

Qy   3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
Db   9260 GACGGATTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 9319

Qy   3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
Db   9320 GATCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCTGGTTCCGCGGGAGGCGGGGCTG 9379

Qy   3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
Db   9380 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTCGAGGCCCCGATTCCATTTGTT 9439

Qy   3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
Db   9440 GAATTCGGCCGATTCCTGAGCTGCCTGAGCAGCGTGGAGCCCGACGGCGCGGCCCTTCCC 9499

Qy   3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
Db   9500 AAATTCCGTTATCCATCGGCTGGCAGCACGTACCCGGTGCAAACCTACGCGTACGCCAAA 9559

Qy   3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
Db   9560 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 9619

Qy   3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
Db   9620 TTGCTGAAGGTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCCGCAAAACTTCGAC 9679

Qy   3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
```

```
Db       9680 GTGTTCGATGAAGCGGCGTTCGGCCTCCTGTTCGTGGGCAGGATCGATGCCATCGAGTCG 9739

Qy       3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
Db       9740 CTGTATGGATCGTTGTCACGAGAATTCTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 9799

Qy       3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
Db       9800 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGTCAATTCGATTTT 9859

Qy       3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
Db       9860 GAACAGGTTCGGCCGGTTCTCGACCTGCGGCATTCGGACGTTTACGTGCACGGCATGCTG 9919

Qy       3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
Db       9920 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 9979

Qy       3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
Db       9980 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGATCAGCACTTCGCCGATATCCTT 10039

Qy       3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
Db      10040 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 10099

Qy       3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
Db      10100 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 10159

Qy       3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
Db      10160 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 10219

Qy       3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
Db      10220 CTCGTTGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTTGGGCTCCAGCAGAGCTTC 10279

Qy       4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
Db      10280 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGTCTGTTGCAGAAGAGG 10339

Qy       4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
Db      10340 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 10399

Qy       4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
Db      10400 TCCGGTTTGCGCCGAGACTCGAAAGATCTAGAGCAGCGGCCGAACATGCAGGACCGAGTG 10459

Qy       4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
Db      10460 GAGGCTCGGCGCAAGGGCAGGAGACGTAGCTAA 10492
```

```
RESULT 9
AR172664
LOCUS       AR172664                71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1..71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others
ORIGIN

  Query Match            97.2%;  Score 4114.6;  DB 6;  Length 71989;
  Best Local Similarity  98.3%;  Pred. No. 0;
  Matches 4159;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          1 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGAT 60
Db       6260 ATGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTGTCAAGCTGGCGGCCGAT 6319

Qy         61 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 120
Db       6320 GGGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGA 6379

Qy        121 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCC 180
Db       6380 ATCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAGTCC 6439

Qy        181 ATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCC 240
Db       6440 ATCGTGCCCGCCCCAGCCGAGCGGCACGTTCCGTTTCCTCTCACAGACATCCAAGGATCC 6499

Qy        241 TACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 300
Db       6500 TACTGGCTGGGTCGGACAGGAGCGTTTACGGTCCCCAGCGGGATCCACGCCTATCGCGAA 6559

Qy        301 TACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 360
Db       6560 TACGACTGTACGGATCTCGACGTGGCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCG 6619

Qy        361 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 420
Db       6620 CGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAA 6679

Qy        421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 480
Db       6680 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCG 6739

Qy        481 AGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 540
Db       6740 AGGCTCGTATCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCG 6799

Qy        541 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATC 600
Db       6800 CTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCAGCAAACCCGTCTCGTGCTCAGTATC 6859

Qy        601 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
Db       6860 GATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGATTGGCTCAGCTTC 6919

Qy        661 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTC 720
Db       6920 TACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGACTATGTGCTC 6979

Qy        721 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 780
Db       6980 GCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGG 7039

Qy        781 CGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTG 840
Db       7040 CGCGTCGCCGAGCTCCCACCTCCGCCGATGCTTCCGATGAAGGCCGATCCATCTACCCTG 7099

Qy        841 AAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTG 900
Db       7100 AGGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGAGTCGATTG 7159

Qy        901 AAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCC 960
Db       7160 AAGCAGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATTCTGGCTGCATTTTCC 7219

Qy        961 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 1020
Db       7220 GAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAAC 7279
```

```
Qy   1021 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 1080
Db   7280 CGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTC 7339

Qy   1081 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 1140
Db   7340 CTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG 7399

Qy   1141 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 1200
Db   7400 CAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCC 7459

Qy   1201 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTT 1260
Db   7460 GCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTC 7519

Qy   1261 AACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGC 1320
Db   7520 AACCAGCAAGTCGTTGGTGTCACCTCGCTGCAGAGGCTCGGCACTCCGGTGTACACCAGC 7579

Qy   1321 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 1380
Db   7580 ACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTC 7639

Qy   1381 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 1440
Db   7640 CTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAA 7699

Qy   1441 GCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCATGGGGTGAACAGGTGCGCTGT 1500
Db   7700 GCGTACGTCGCTTTTCTCCGGCGGCTCACTGAGGAACCATGGAGTGAACAGATGCGCTGT 7759

Qy   1501 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTG 1560
Db   7760 TCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGAGACCAACTCGCTGCTG 7819

Qy   1561 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTC 1620
Db   7820 AGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCTATGCAGCTC 7879

Qy   1621 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 1680
Db   7880 GCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGA 7939

Qy   1681 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 1740
Db   7940 CTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATG 7999

Qy   1741 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 1800
Db   8000 GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTAC 8059

Qy   1801 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 1860
Db   8060 GTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAG 8119

Qy   1861 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 1920
Db   8120 GTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATC 8179

Qy   1921 CAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATG 1980
Db   8180 CAGCGGCTGCTCGTGAGCGATGCCGGCGTCGAAGGCGACGGCGACCAGCTTCCGATGATG 8239

Qy   1981 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 2040
Db   8240 CCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTG 8299

Qy   2041 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 2100
Db   8300 CCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAG 8359

Qy   2101 CGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 2160
Db   8360 CGCTTCGAAATAGGGCCCGGAGACAGAGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTC 8419

Qy   2161 TCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 2220
Db   8420 TCGGTCTACGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCG 8479

Qy   2221 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTG 2280
Db   8480 TCCAAGCTGCGCGATCCGGCGCATTGGGCAGCGTTGATCGAACGAGAGAAGGTGACGGTG 8539

Qy   2281 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGAT 2340
Db   8540 TGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTCCGAGGGTCGCCCCGAT 8599

Qy   2341 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 2400
Db   8600 TCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTG 8659

Qy   2401 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 2460
Db   8660 CCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACC 8719

Qy   2461 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGC 2520
Db   8720 GAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGATCCATCGTGGGCGAGC 8779

Qy   2521 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 2580
Db   8780 ATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAA 8839

Qy   2581 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGC 2640
Db   8840 CCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGACTGGCACTGGGC 8899

Qy   2641 TACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAG 2700
Db   8900 TACTGGCGCGATGAAGAGAAGACGCGCAACAGCTTCCTCGTGCACCCCGAGACCGGGGAG 8959

Qy   2701 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 2760
Db   8960 CGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATG 9019

Qy   2761 GGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 2820
Db   9020 GGGCGGGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAG 9079

Qy   2821 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 2880
Db   9080 GAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGAC 9139

Qy   2881 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCC 2940
Db   9140 GCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAAGGCACACGGAGACGCGCTGCC 9199

Qy   2941 GAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCGCCGAAGCG 3000
Db   9200 GAGCAGGACGCGAGCCTCAAGACCGAGCGGGTCGACGCGAGAGCACACGCCGCCAAAGCG 9259

Qy   3001 GACGGCTTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 3060
Db   9260 GACGGATTGAGCGACGGCGAGAGGGTGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGG 9319

Qy   3061 GACCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTG 3120
Db   9320 GATCTGGACGGAAAGCCCGTCGTCGATCTGACCGGGCTGGTTCCGCGGGAGGCGGGGCTG 9379

Qy   3121 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTT 3180
Db   9380 GACGTCTACGCGCGTCGCCGTAGCGTCCGAACGTTCCTCGAGGCCCCGATTCCATTTGTT 9439

Qy   3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCC 3240
```

```
               || || || ||||||||||||||| ||||||||||||||||||||||||| ||||||||
Db        9440 GAATTCGGCCGATTCCTGAGCTGCCTGAGCAGCGTGGAGCCCGACGGCGCGGCCCTTCCC 9499

Qy        3241 AAATTCCGTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAA 3300
               |||||||||||||||||||| |||||||||||||||||||||||||||||||| | ||||
Db        9500 AAATTCCGTTATCCATCGGCTGGCAGCACGTACCCGGTGCAAACCTACGCGTACGCCAAA 9559

Qy        3301 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 3360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9560 TCCGGCCGCATCGAGGGCGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGT 9619

Qy        3361 TTGCTGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGAC 3420
               ||||||||| |||||||||||||||||||||||||||||||||||| |||||||||||||
Db        9620 TTGCTGAAGGTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCCGCAAAACTTCGAC 9679

Qy        3421 GTGTTCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCG 3480
               |||||||||||||||||||||  |||||||||||||||||||||||| ||||||||||||
Db        9680 GTGTTCGATGAAGCGGCGTTCGGCCTCCTGTTCGTGGGCAGGATCGATGCCATCGAGTCG 9739

Qy        3481 CTGTATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 3540
               ||||||||||||| ||| |||||||| |||||||||||||||||||||||||||||||||
Db        9740 CTGTATGGATCGTTGTCACGAGAATTCTGCCTGCTGGAGGCCGGATATATGGCGCAGCTC 9799

Qy        3541 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTT 3600
               ||||||||||||||||||||||||||||||||||||||||||||||| |||||| |||||
Db        9800 CTGATGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGTCAATTCGATTTT 9859

Qy        3601 GAACAGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTG 3660
               ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        9860 GAACAGGTTCGGCCGGTTCTCGACCTGCGGCATTCGGACGTTTACGTGCACGGCATGCTG 9919

Qy        3661 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 3720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9920 GGCGGGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCG 9979

Qy        3721 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTT 3780
               |||||||||||||||||||||||||||||||||||||| ||||||||||||||||| |||
Db        9980 AGGCGCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGATCAGCACTTCGCCGATATCCTT 10039

Qy        3781 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 3840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10040 CGCGACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTC 10099

Qy        3841 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 3900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10100 GATGCGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAG 10159

Qy        3901 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 3960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10160 GATACCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATC 10219

Qy        3961 CTCGTCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTC 4020
               ||||| ||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       10220 CTCGTTGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTTGGGCTCCAGCAGAGCTTC 10279

Qy        4021 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGG 4080
               ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       10280 GTCGATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGTCTGTTGCAGAAGAGG 10339

Qy        4081 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10340 CTGGATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 10399

Qy        4141 TCCGGTTTGCGCCGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGCAGGACCGAGTG 4200
               |||||||||||||||||||||| ||||||||| |||||||||||||||||||||||||||
Db       10400 TCCGGTTTGCGCCGAGACTCGAAAGATCTAGAGCAGCGGCCGAACATGCAGGACCGAGTG 10459

Qy        4201 GAGGTTCGGCGCAAGGGCAGGAGACGTAGCTAA 4233
               |||| ||||||||||||||||||||||||||||
Db       10460 GAGGCTCGGCGCAAGGGCAGGAGACGTAGCTAA 10492
```

RESULT 10

```
AF188287
LOCUS       AF188287               42603 bp    DNA     linear   BCT 12-JAN-2000
DEFINITION  Stigmatella aurantiaca myxothiazol gene cluster, complete sequence.
ACCESSION   AF188287
VERSION     AF188287.1  GI:6635393
KEYWORDS    .
SOURCE      Stigmatella aurantiaca
  ORGANISM  Stigmatella aurantiaca
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE   1  (bases 1 to 42603)
  AUTHORS   Silakowski,B., Schairer,H.U., Ehret,H., Kunze,B., Weinig,S.,
            Nordsiek,G., Brandt,P., Blocker,H., Hofle,G., Beyer,S. and
            Muller,R.
  TITLE     New lessons for combinatorial biosynthesis from myxobacteria. The
            myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca
            DW4/3-1
  JOURNAL   J. Biol. Chem. 274 (52), 37391-37399 (1999)
  MEDLINE   20069734
   PUBMED   10601310
REFERENCE   2  (bases 1 to 42603)
  AUTHORS   Silakowski,B., Schairer,H.U., Ehret,H., Kunze,B., Weinig,S.,
            Nordsiek,G., Brandt,P., Bloecker,H., Hoefle,G., Beyer,S. and
            Mueller,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-1999) NBI/MX, GBF, Mascheroder Weg 1,
            Braunschweig 38124, Germany
FEATURES             Location/Qualifiers
     source          1..42603
                     /organism="Stigmatella aurantiaca"
                     /mol_type="genomic DNA"
                     /strain="DW4/3-1"
                     /db_xref="taxon:41"
     gene            1..834
                     /gene="mtaA"
     CDS             1..834
                     /gene="mtaA"
                     /note="Ppant transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="MtaA"
                     /protein_id="AAF19809.1"
                     /db_xref="GI:6635394"
                     /translation="MPTSSPALPLLKLPPDEVHVWIVEPERITEPGLLESYRALLDPG
                     ERDKQQRFYFERHRLQYLVSHALVRLTLSRYAPVAPEAWSFSANQYGRPEIRGEEKPW
                     LRFNLSHTDGMALCAVARDVDVGADVEDTERRGETVEIADSFFAPAEVASLRALPVSG
                     QRERFFDYWTLKEAYIKARGMGLSLPLDQFAFEVSQGLSTRISFDPRLVDEPSQWQFV
                     RPRPSQRHAAALAVRRPSEAPLTVRFQRTVPLQDDAPAEYLSRERIQPLRLRMPGVGG
                     G"
     gene            1388..13399
                     /gene="mtaB"
                     /note="PKS"
     CDS             1388..13399
                     /gene="mtaB"
                     /note="PKS; MtaB loading domain and module 1: polyketide
                     synthase domains: ACP, KS, AT, AT, DH, ER, KR, ACP; MtaB
                     module 2: polyketide synthase domains: KS, AT, DH, KR,
                     ACP"
                     /codon_start=1
                     /transl_table=11
                     /product="MtaB"
                     /protein_id="AAF19810.1"
                     /db_xref="GI:6635395"
                     /translation="MSTEHGVGGVGRSTGQLRDWLVQQVASQVGLEAHTVRITEPFFS
                     LGVGSVFATRLIAELARWLERPLSPTLVWEHPTIDKFARHLSGAEPPRAEMHAPGPEA
                     AVDGAEPIAIVGLACRFPKAPDAKAFWRVLQEGRDCVTEVPSNRWDASLYYDPAPGTP
                     GKIVTRYGAFLDSVDRFDPVFFGISPREAGQLDPQQRLMLELSWEALEDAGIAPRVLR
                     GSRTGVFTGMVWRDYADLQTIHNAAPTLHTCTGYAGGIIANRISYALGLEGPSLAIDT
                     ACSSSLVAIHEACRSLRSGESSLALAGGVQLNLSPRTFMAVSAFGALAPDGRCRTFDA
                     RASGYVRGEGGGVLVLKPLSRALADGDAIYCLIRGSAVNNDGASNGLTAPNPKAQERL
                     LRDAYAVAGMDPAQVHYVELHGTGTPLGDPIEAGALGAVLGAAREAERPLHVGSVKTN
                     IGHLEAAAGIASVIKVALSIRHQQLPASLHFEQPNPRIPFDDLRLRVQTQLAPWPAEK
                     DPAVAGISAFGFGGTNCHVVVEEHRGGSRTAAVPVSTPHAQSPRHPKTAFLFAGAGSA
```

```
                     WWGMGRELLGAEPVFRARFEECDAAFRPLGGWSLVEALLAPRVRSRFEQGERQLPLVL
                     AFQISLAALWRSWGIEPSATVGYSVGEWAAAQVAGILTIEEVFALAYHYIQVQKGVAG
                     GVGMAVVGLSGAEVAERIAPYGGRISIASENGPRSTGLGGEAAALQDFTASMTAAGIF
                     ARLVEIDFYAHIPQVEALKPLFLQSCPVLRPSPGRVPMVSTVSGAFVDGGTLGAEYWA
                     SNFREPVRFATAMRTLLEGGYDVVLDVNPHPIHEHSVLENVRAGTASAEFIASMRRGE
                     PSVEGLHRALTQLSQAGGPVRWERVVPGTGAGSSAPVRTELFVASARTEQALREQADR
                     IAEQLRRDAASSLTDLCYTAGARRSHHEHRLALVAETRAALEEGLAAFARGEIRPNPS
                     QGMVRPGEAPRLGFVFSGQGSFPLGSSVQLFEEEPSFRKAFERSERLFRAQTGASLIE
                     ALHAERESQHPPRTEIEQPLLFAFQVALAALWESWGIRPDAVVGHSVGEISAAHVAGV
                     LSLEEALSLVVSRSRLMGRAAGQGGMATVEASPARIEPLLGKCGTDLAIAAVNGPENV
                     VVSGALAALASLRDELTGQGLRVRVLNVNYAFHSAQMAPYLGDLDPVLRELKPKPPRL
                     PIYSTVHGRLASASDFGTDHWRANVRERVLFAPAVEAMRADGHGAFVEISPHPVLLQP
                     LSRILDASPASGASALFSTRRDRAPRAALLGNLGTLYTLGHEPSWKGVFPDLGRTVEL
                     PSYPWQRQRCWFDAAPQPLAEPVHPLLGRRVDAPPGQTVFEATFRSDRLQLLSDHRLY
                     GTVVVAGASMLGMMLVAARQLWGEAPCALQDVLFLQPLVLGDTEDRLVQVSIATDGAF
                     RILSWAGQGGEGTSSHWIEHATGRLERGGAEETEAGREVLSEVEARCRPLASAESFYE
                     GIARQGLGLGRQFRWLEALWQGDREAFGRLRVPETDDGAEDYRFHPGLIDSCFQVLGA
                     VAGFASIEGLAFVPIGVGRFALRGEARPPRFCHLVLRRGSGKANEASSVDLRLLDESG
                     LVVAEVSGLVIRRAPREVMLRRDVTAPPLVLRTEWQPQMLSGASQAGPEAWLLLADRG
                     GLAARLAEQLSSRGSRVEVLAAEENAQSMLGRIEAIDASGLACAGVVDLRGLDEGSSS
                     DAVQGAVKRVQQLVTRRAAEALRLVLVTRGTQVTGSQDGSIAFEQAPLWGLGNVIALE
                     HPELRCQRIDLDPSPGAPDAAQLAAELAAPGREDRVAFREGRRLVARLVRTKLPPAVR
                     ALPLEGDATYLIVGGLGGIGLELARWMVGRGARHLVLMGRSAPSAHAQETLAALRDAG
                     ATISVALGNVAEREALARVLGDIRAAMPPLRGVMHAAGVLDDGLLVRQSEERLARVLA
                     PKIQGAMNLHRETVGDRLDFFVLFSSVSSVFGTSGQAGYASANAFLDALALHRRAQGL
                     TALSINWGPWADVGMAARLPPEAGKSWSSFGLGLLAPSAGLAALERLLSSEAAQVVVA
                     GASEAEAKGDFAPFFDALSSRKPEASRPASVGGALVGASPRELRARLLSLLQVQAATV
                     LGLPPSAPIDPRRPLQELGLDSLMALQFRNSLAASLGRSLPATLLFDHPTLEKLFEFL
                     ARDLTPASEASPAKSRRPVKLPARDAEGAEPIAVVGMACRLPGSVRSPEEFWRLLANG
                     VNAVREVPKDRWSLDTWYDPDPETPGKMYSRHGGFLDEVDPFDASFFGISAPEARSMD
                     PQQRILLELGWEALESAGIAVDQLKGSPTGVFLGICLSDYALLELNSPDPTGINAYSG
                     SGGVLSVAAGRIAYALGLEGPTLAVDTACSSSLVAAHLACQSLRNGECDLALVGGSNL
                     LLSPRMSVYFSKLKALSPDGACKAFDSSANGYVRGEGAGVLVLKRLSEALADGDPILS
                     VMRGTAVNQDGRSNGLTAPSRAAQERVIESALRQGGVSPLDVSYVEAHGTGTSLGDPI
                     EVQALAAVLGKDRPRSAPLGIGSVKTNVGHLEGAAGVTSLIKAILALRHRELPRSLHF
                     KTPSPHIPWDELPVEVISEHRPWRVPEGKRRILGVSSFGFSGTNAHVILEEAPEAGPR
                     SKAPSIPERPELLVLSARAPQALQEAAVRFSAQLSGRAGGDAAMLHDVCYSANCRRSH
                     HEHRLAVVARSRTEAAEALEAFLRGLAHPDVQAGYALPGAVPRVVFLFTGQGGQWPGT
                     GRALLEEEPVFRSALQAADEVLKGLAGWSIIEELQRETPRFDQAAISQPAHFALQSAL
                     LALLRSWGIEPDAVAGSSIGEVAAAYAAGVLTLEEAMRITVARSRILHAVAGKGAMAL
                     VELSEEEATTEVAAYAQRLSVAAINSQNSTLLSGEAGAMDELLGALTARGVSCRRIGL
                     DYASHSPAVEPYIDELKQALQGLRPQPAAIPIFSTVTGRLAEAGDFNSDYWARNLRQP
                     VRLAPVLTTLAREDYTLFVELGPNAVMARPVQQCFQAVGKSATLLSTLRRGESERHTL
                     RSLVGGLFIAGVTPAWNRLFPEGGSLVSLPSYPWQKERYWIASPTSARLQLPAPSQGV
                     LHGRRVRSPLAESLFETDLSVEKMPFLREHLLHGALVVAGAAHVALALSALRELRGTA
                     PVRLADIVFPRPLVLKEGVENPVSVVVGREQGGRSALRIQSPQGSEPEWVTHAEGWVL
                     PASSQVEGPAQVALEELRARCPNGLEVSPVYKAMDAEGIGLGRRFQWVEQLWQGDGEA
                     LGRMRAAESADQHPSAPLHPGLIDSIFHVLRAVPGIGRGDGVPYIPIGIEGLCFLAPP
                     RGVLWAHVKLRPEGSRDGETLSADFRLFEESGKTVAWIDGLIVKRAPREAFQRHEAQD
                     LSEKLFQVRWQSMPLAPAVPERGPWLLVADRGPAGEELAARLKREGQVCHVLAAPESA
                     VTGEALLQRVEEALALEAYRGVVYLSALEAEQESPQGKSQAALEGVSGAVQVLQAIAK
                     LERGTPPRLWLVTRGAQALEESPRPLALAQAPLWGLGLAVRDEHPELRCTLIDLDPDP
                     AADGVGCLFQELLTPDREDRIAWRGGARFAQRLSFLEAVPEPGTAPISPDGTYLITGG
                     LGGIGLRCARWLVDAGARHLALLGRSGPSEAARSVLGALEAEGVTVRVLQADVAVRED
                     VSKVLSALRVDMPPLRGVLHAAGLLDDALLVRQDRDRIQRVLQPKVAGALNLDLETRG
                     DPLDFFVLFSSIAAMLGGEGQGNYSAANAFLDALAFARRAQGLPAHSLAWGPWAEVGM
                     VAARSERSAHLAALGIESIPPDEGVAILGRLLRAASAHVGIFQIDWTRWQQRHAEGHA
                     PLLAGILKQQAAAPSQDSSEEAPDIQELIRATPEERQKRAEAFLRREVTRILDLPRPL
                     ADGRLPLVRFGFDSLMSLKLKARLESTLGLSISAAKLLSGLSLDKLVALVLERIETLQ
                     PTPVSTDDQMEDFRF"
     gene            13537. .17409
                     /gene="mtaC"
     CDS             13537. .17409
                     /gene="mtaC"
                     /note="peptide synthetase; peptide synthetase domains:
                     Heterocyclisation, adenylation, PCP note: motif with weak
                     homology to putative NAD binding proteins from Pyrococcus
                     species located behind PCP; this motif may be involved in
                     thiazole formation from the proposed thiazoline
                     intermediate of the biosynthesis of myxothiazol"
                     /codon_start=1
                     /transl_table=11
                     /product="MtaC"
                     /protein_id="AAF19811.1"
                     /db_xref="GI:6635396"
                     /translation="MIELLRPKEAVPDAARQFEPFPLTDIQETYFVGRGTDFGISGMS
                     CHLYHELDTRGMDVARLSRAWQRLIEQHVMLRSVVLPTGEQRVLESVPPFSIPVLDLR
                     GQAPALVETKLAEIRQQLSTRSTPPGQWPTFELRATLLDEGRTRIHFDFDAITADASA
                     VLTLLEEWREFYLSPDRPLKQVPVSFRDHVLAEVAARDTDSFRQAEAYWNARLNDMPD
                     APALPVATEPERLMRPSFRRLSGRLDAARWARLKERAANAGLTASGVVCAAFAEVLGA
                     WSESPRFTLNLTLFRRLPIHPEVDRLLGDFTTNVLLEVDRSGESFTARAVRLRDQLAE
                     DLEHTHFSGVRVMRQRAQMKKGNGAIMPVVFTSLLGHRSRVPGGGSPFSWLGEVVYAI
                     TQTPQMWLDHQVREEAGALYCHWDSPEGLFPEGLLEGAFEAYTDLLSRLADDNALWEE
                     GTVVRLPEAQRARRETFNATTASIPEARLDGLFLAQAEQGPQRMAILEPALRMSYGEL
                     LGQAGSLAAHLAELGARPNELTAVVLEKGWRQIVAVTAIHLAGSAYLPIDPALPEERR
                     RLLLEEGQVKVVLTESHLAERLSWPASVRVVAIADTARGAPPRLPARAASDMAYCIYT
                     SGSTGRPKGVKIEHRAAANTILDINERFGVGPADRVFALSSLGFDLSVYDIFGSLAAG
                     AAIVLPQPEVTRDPAHWLQVLRQQGVTIWNSVPTLMEMLVDQVEARGELLPETLRLVM
                     LSGDWIPVTLPDRIRRLSRDIRVISLGGATEASIWSIIHPIGRVDPAARSIPYGRPML
                     NQRFYVLDEKLAPRPDHVPGDLYIGGVGLAQGYFRDEETTRARFFNHPRTGERLYRTG
                     DLGRFLPEGEIEFLGRKDFQVKIAGHRIELGEIEAVLQRHPSIREAVVAAPGERTSRR
                     LVAYLVPVEGQEAPSDEVLRSFLAAALPQYMVPAVFVRLEHLPLSSNGKVDRKALPEP
                     VLQPSRASGLTDARGGKILSAIMRLVSEVLKRPSIEPEAGLLQLGATSVELIKLSMML
                     EQEFGSRPTITELLSLRNASEIAAYYVERQPAEEARKSDEAELVPAGIRRFDAGAVTA
                     QLEAPPEDEARRQCYFKRVSHRTFSAEPVEARVLARLLSCLSPLKVEGQLKYQYGSAG
                     GIYGVQTYLSIATGRVRGLDAGAYYYDPLRHTLVQLGGASALPAALDSSANAAMADSA
                     AFSLYLVGDRRVISQRYGERWRDLALIEAGLMAQLLETRAAESDLGLCQLEELRFEGL
                     SKALRLEEEHMYLHGLVGGGLAWEEGAL"
     gene            17406. .27281
                     /gene="mtaD"
     CDS             17406. .27281
                     /gene="mtaD"
                     /note="PKS/NRPS; MtaD module 1: peptide synthetase
                     domains: heterocyclisation, adenylation, PCP note: motif
                     with weak homology to putative NAD binding proteins from
                     Pyrococcus species located between adenylation core
                     regions A9 and A10; this motif may be involved in thiazole
                     formation from the proposed thiazoline intermediate of the
                     biosynthesis of myxothiazol; similar to the motif found in
                     MtaC; MtaD module 2: polyketide synthase domains: KS, AT,
                     DH, KR, ACP"
                     /codon_start=1
                     /transl_table=11
                     /product="MtaD"
                     /protein_id="AAF19812.1"
                     /db_xref="GI:6635397"
                     /translation="MSTSEFIVYLQQKGIELWIEGDTLRFRAPPGHLTPELRTALKDK
                     KEQLLEHLRSASRSANLQTVASPPDAASRFAPFPLTDIQNAYWVGRQGAFDAGGVAAH
                     SYLELAFDTLDLPRLESILQRLIEYHDMLRMVVLPTGEQVVLEKVPPFTLTAYDLAGA
                     PASQIEAHFEAIRGELSHQVLPTDRFPLFDVRASKLAGGRVHLHISFDLLTADAFSVQ
                     ILIEQCASLYLNPEAQLVPFKRTFRDYCVLAAQRADPSSRAYQNSLAYWRNRLEKLPN
                     APDLPLVNDTDVKGQHQFVRRKAGLDQASWRAFRNHAKAAGITASMALCAAYGETLRA
                     YSRSNRLTLNLTLFNRLPLIEDVERIVGDFTSGILLELDGSKPESFAARARRLQGQFI
                     NDLEHSAVSSVKVLREAIRLGRWDAGTAMPYVFTSLISETGRSLSVSEGVKIVDVISQ
                     TPQVWLDHQVFELDGGIYYSWDSVDALFPAGLIDAVFESYGRLLHRLSSDASAWDALS
                     RQPLLPAQLASRAEYNATERPVSQERIESLFLRQAERQPGAVAVIAGEQVLSYGKLER
                     RSARIAAWLIAQGVKPDQLVAIVAEKGVEQVVAALAILRAGAAYLPLSPSLPPERLHA
                     LLEEAQADVVLTQSALESSLRWPEGPRRLSVDRDEFLEAPKGALPPVQGNNLAYVIYT

  Query Match             22.3%;  Score 944.8;  DB 1;  Length 42603;
  Best Local Similarity   54.1%;  Pred. No. 3.5e-139;
  Matches 2269;  Conservative    0;  Mismatches 1807;  Indels  120;  Gaps   12;

Qy         14 AGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGATGGGGAGCGCCTCC 73
              || |  | .| | ||| ||||  |||| ||| | |||   | ||| |||||     || |
Db      17419 AGTTCATCGTCTATCTCCAGCAGAAGGGCATCGAACTGTGGATCGAAGGGGACACGCTGC 17478

Qy         74 AGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGAATCTCCGAGCACA 133
                | | || |||      | |||| |  ||  | |||| | |    | |    || | |
Db      17479 GATTCCGCGCGCCCCCTGGCCACCTCACGCCCGAGCTGCGCACGGCACTGAAGGACAAGA 17538

Qy        134 AAAGCACGATCCTGACGATGCTCCGTCA---------------GAGACTCCCCGCAGAAT 178
              |      | |||||  |    || |||                 || ||  |    | |
Db      17539 AGGAGCAGCTCCTGGAGCATCTGCGTTCCGCGAGCCGCTCCGCGAATCTTCAGACCGTGG 17598

Qy        179 CCATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAAT 238
```

```
Db      17599 CGAGCCCGCCGGATGCGGCGTCACGCTTCGCCCCTTTCCCGCTCACCGACATCCAGAATG 17658

Qy        239 CCTACTGGCTGGGCCGGACAGGAGCGTT---TACGGTCCCCAGCGGGATCCACGCCTATC 295
Db      17659 CCTACTGGGTGGGAAGGCAAGGTGCTTTCGATGCGGGCGGCGTCGCAGCGCACAGTTACC 17718

Qy        296 GCGAATACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCG 355
Db      17719 TCGAGCTCGCGTTCGACACGCTCGATCTGCCGCGCTTGGAGAGCATCCTTCAGCGGCTCA 17778

Qy        356 TCGCGCGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGC 415
Db      17779 TCGAGTACCACGACATGCTCCGCATGGTGGTGCTGCCGACGGGCGAGCAGGTGGTGCTGG 17838

Qy        416 CTAAAGTCGACGCCGACATCGAGATC---ATCGATCTGCGCGGGCTCGACCGGAGCACAC 472
Db      17839 AGAAGGTGCCGCCGTTCACCCTCACCGCGTATGACCTTGCTGGGGCGCCCGCCAGTCAGA 17898

Qy        473 GGGAAGCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGC 532
Db      17899 TCGAGGCGCATTTCGAGGCCATCCGCGGGGAACTCTCCCACCAGGTGCTTCCGACGGACC 17958

Qy        533 GCCCTCCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGC 592
Db      17959 GGTTCCCCCTGTTCGATGTCCGGGCCTCGAAGTTGGCCGGTGGACGGGTTCACCTCCACA 18018

Qy        593 TCAGTATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGC 652
Db      18019 TCAGCTTCGATCTCCTCACGGCCGATGCGTTCAGCGTCCAGATCCTGATCGAACAGTGCG 18078

Qy        653 TCAGCTTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATT 712
Db      18079 CTTCGCTGTACCTGAATCCGGAAGCGCAACTGGTGCCCTTCAAGCGGACCTTCCGGGATT 18138

Qy        713 A------TGTACTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGG 766
Db      18139 ACTGCGTCCTGGCCGCCCAGCGCGCTGACCCGAGTTCCCGGGCCTACCAGAACTCGCTTG 18198

Qy        767 ATTACTGGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCG 826
Db      18199 CGTACTGGCGAAACCGCCTCGAGAAGTTGCCCAACGCCCCCGATCTGCCGCTCGTGAACG 18258

Qy        827 ATCCATCTACCCTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACT 886
Db      18259 ACACGGACGTGAAGGGGCAGCACCAGTTCGTGCGCCGGAAGGCAGGGTTGGATCAGGCCT 18318

Qy        887 CCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCC 946
Db      18319 CCTGGCGCGCATTCCGCAATCATGCCAAGGCCGCGGGGATCACGGCGTCCATGGCGCTGT 18378

Qy        947 TGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACA 1006
Db      18379 GCGCCGCCTATGGGGAAACGCTGCGGGCCTACAGCCGGAGCAACCGGCTCACGCTCAACC 18438

Qy       1007 TAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCA 1066
Db      18439 TGACCCTCTTCAACCGCCTGCCCTTGATCGAGGACGTGGAGCGGATCGTCGGGGACTTCA 18498

Qy       1067 CGTCGATGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTA 1126
Db      18499 CCTCGGGAATCCTGCTCGAACTGGATGGCTCGAAGCCCGAGTCCTTCGCCGCCCGGGCGC 18558

Qy       1127 AGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGG 1186
Db      18559 GGAGGCTCCAGGGGCAGTTCATCAATGATCTGGAGCACTCGGCTGTTTCGAGCGTGAAGG 18618

Qy       1187 TCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGC 1246
Db      18619 TGCTGCGGGAAGCGATCCGCCTTGGCCGCTGGGATGCCGGGACCGCGATGCCCTATGTCT 18678

Qy       1247 TCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTC 1306
Db      18679 TCACGAGCCTCATCTCGGAGACCGGCCGCTCATTGAGCGTCTCCGAGGG---CGTGAAGA 18735

Qy       1307 CGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACG 1366
Db      18736 TCGTCGACGTCATCTCCCAGACGCCCCAGGTGTGGCTCGACCACCAGGTTTCGAGCTGG 18795

Qy       1367 ATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGG 1426
Db      18796 ATGGCGGCATCTATTACAGCTGGGATTCCGTCGATGCCCTCTTTCCCGCGGGCCTGATCG 18855

Qy       1427 ACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGAACCA------T 1480
Db      18856 ACGCTGTCTTCGAGAGCTACGGAAGGCTCCTCCACCGTCTCTCCAGCGATGCCTCGGCGT 18915

Qy       1481 GGGGTGAACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAA 1540
Db      18916 GGGATGCACTCTCGAGGCAGCCGCTCTTGCCGGCCCAGTTGGCAAGCCGTGCCGAGTACA 18975

Qy       1541 ACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCG 1600
Db      18976 ACGCCACCGAGCGTCCCGTGTCGCAGGAGCGGATCGAATCGCTCTTCCTTCGTCAGGCCG 19035

Qy       1601 AGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGC 1660
Db      19036 AGCGTCAACCCGGGGCCGTCGCGGTTATCGCGGGAGAGCAGGTTCTCTCCTATGGCAAGC 19095

Qy       1661 TTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACA 1720
Db      19096 TCGAGCGCCGGTCGGCTCGGATCGCAGCGTGGCTCATCGCCCAGGGTGTCAAGCCGGACC 19155

Qy       1721 CATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGC 1780
Db      19156 AGCTCGTGGCGATCGTTGCCGAGAAGGGGGTGGAGCAGGTGGTCGCCGCGCTCGCCATCC 19215

Qy       1781 TCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACT 1840
Db      19216 TTCGCGCTGGAGCCGCCTACCTGCCGTTGTCTCCATCCCTGCCCCCCGAGCGGCTCCACG 19275

Qy       1841 ACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAAC 1900
Db      19276 CCTTGCTGGAAGAGGCGCAAGCCGACGTGGTGCTCACCCAGTCCGCCCTGGAGTCGTCGC 19335

Qy       1901 TGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACG 1960
Db      19336 TGCGATGGCCCGAGGGGCCTCGTCGGCTTTCGGT------GGATCGTGACGAGTTTCTCG 19389

Qy       1961 GCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACA 2020
Db      19390 AAGCCCCGAAGGGGGCGCTGCCGCCGGTTCAGGGCAACAATCTGGCCTACGTCATCTACA 19449

Qy       2021 CCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACA 2080
Db      19450 CCTCGGGCTCGACGGGCCGTCCCAAGGGCGTGATGATCGACCACCGGGGAGCGGTGAACA 19509

Qy       2081 CCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCT 2140
Db      19510 CCCTCCTCGACATGAACGAGCGCTTCAATGTGGGGCCCCGGGACCGGGTGCTGGCGCTTT 19569

Qy       2141 CCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTA 2200
Db      19570 CGTCACTCAGCTTCGACTTGTCTGTCTACGACCTCTTCGGAGTTCTTGCCGCGGGCGGTG 19629

Qy       2201 CGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCG 2260
Db      19630 CGATCGTGATGCCGGAGCCCGGCACGTCGCGGGATCCCGGACGGTGGCAGGTGCTGCTGG 19689

Qy       2261 AACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGC 2320
Db      19690 AGAAGACGGGGGTGACGATCTGGAACTCGGTCCCCGCGTTGATGGACATGCTCGTCGAGT 19749

Qy       2321 ATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCG 2380
Db      19750 TCTCGGAGGGAAGCGGTCTGCGCTTGCCGGACTCGCTGCGCCTGGTCCTCATGAGCGGTG 19809
```

```
Qy     2381 ACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGA 2440
Db    19810 ACTGGATTCCGGTCACGCTGCCCGGCCGGATCCGTGCCCTGTCCAAGAACGCCGAGCTGG 19869

Qy     2441 TCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACG 2500
Db    19870 TCAGCCTCGGTGGCGCGACCGAGGCCTCCATCTGGTCCATTCTCTACCGGATTGGCGATG 19929

Qy     2501 TCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACG 2560
Db    19930 TGGGGGCCGGCTGGCGCAGCATTCCGTACGGTCGGCCCATGGTGAATCAGCGCTTCTATG 19989

Qy     2561 TGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCG 2620
Db    19990 TGCTGGATGAGGCGCTGGAGCCCTGCCCGGACTGGGTCGCGGGCCAGATGTACATTGGCG 20049

Qy     2621 GGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCG 2680
Db    20050 GCATTGGCCTCTCGCTCGGCTATTACCGGGATCCGGTTCGTACCGCCGAGCGCTTCATCG 20109

Qy     2681 TGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCG 2740
Db    20110 TCCATCCCAAGACCGGGGAGCGCCTGTACGCAACGGGTGACCTGGGCCGGTTCATGCCGG 20169

Qy     2741 ATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCG 2800
Db    20170 ACGGGAACATCGAGTTCCTCGGACGGGAGGACTTCCAGGTCAAGATCCAGGGCTATCGCA 20229

Qy     2801 TTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGA 2860
Db    20230 TCGAGCTCGGGGAGATCGAGGCCGCGCTCGACTCGCATCCAGCGGTGCGCTCCTCGGTGG 20289

Qy     2861 TTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGG 2920
Db    20290 TCAATGCCGTTGGAAAGCCCGGGGGACTCGGCGCCTCGTGGCCTATGTCGTCCCTGGGG 20349

Qy     2921 GCACACGGAGACGCGCTGCCGAGCAGGACGCGAGCCTCAAGACCGAGC---GGATCGACG 2977
Db    20350 AATTGACGTCGAATGCTCCCCCGGCAAGCGCGCCCTCGGCGAGCACGCTCTCCTTGGAGT 20409

Qy     2978 CGAGAGCACACGCCGCCGAAGCGGACGGCTTGAGCGACGGCGAGAGGG------------ 3025
Db    20410 CGGCTCCGAAGCTTGCCGACGAGTACAGCGCTCTGGAGAGCCAGGGGGCTGGGCTGATCA 20469

Qy     3026 ------------------TGCAGTTCAAGCTCGCTCGACACGGACTCCGGAGGGACCTGG 3067
Db    20470 CCGATACGGTGGAACGCCTGGAGTTCAAGCTGAAGCAACACAACCTGCGCCCGGATGCAG 20529

Qy     3068 ACGGAAAGCCCGTCGTCGATCTGACCGGGCAGGATCCGCGGGAGGCGGGGCTGGACGTCT 3127
Db    20530 GCAGCCGGGCTCAAGTTGCGCTCAAGAAGCCGGAGTTCGATGAGGCCTTCCGCAAGAGGT 20589

Qy     3128 ACGCGCGTCGCCGTAGCGTCCGAACGTTCCTTGAGGCCCCGATTCCGTTTGTTGAGTTTG 3187
Db    20590 ACATGGAGCGGCAGAGCGCCCAGTCGTTCCTCTCCGAGCCCGTCTCTCTGGAGCGGCTCA 20649

Qy     3188 GTCGATTCCTGAGCTGCTTGAGCAGCGTGGAGCCCGACGGCGCGACCCTTCCCAAATTCC 3247
Db    20650 GCGAGTTGCTGAGCTGTCTGATGCAGATCGCTCTCGAAGACTCCATGATGCCCAAGTACC 20709

Qy     3248 GTTATCCATCGGCGGGCAGCACGTACCCGGTGCAAACCTACGCGTATGTCAAATCCGGCC 3307
Db    20710 GGTATGCATCGGCCGGTGGGCTCTATCCGGTGCAGGTCTACCTGCACGTCAAGCCAGGGC 20769

Qy     3308 GCATCGAGGG---CGTGGACGAGGGCTTCTATTATTACCACCCGTTCGAGCACCGTTTGC 3364
Db    20770 GTGTCGGGGGGCTCGCGGGCGGGACCTACTATTATCACCCGAAGCGTCACGAGCTCGTGC 20829

Qy     3365 TGAAGCTCTCCGATCACGGGATCGAGCGCGGAGCGCACGTTCGGCAAAACTTCGACGTGT 3424
Db    20830 TGCTGACCGCGGATGCGGCCATGGACAGGAGCCAACACGCGTCCCGGAACCGGCCCGTCT 20889

Qy     3425 TCGATGAAGCGGCGTTCAACCTCCTGTTCGTGGGCAGGATCGACGCCATCGAGTCGCTGT 3484
Db    20890 TCGATGCCGCCGCGTTCTCCGTTTTCCTCGTTGGAAAGCTGAGTGCCATCGCGCCGCTGT 20949

Qy     3485 ATGGATCGTCGTCGCGAGAATTTTGCCTGCTGGAGGCCGGATATATGGCGCAGCTCCTGA 3544
Db    20950 ACGGGAGCATGGCGCGGGATTTCTGCATGCTGGAGGCGGGCTACATGGCTCAGCTCCTCA 21009

Qy     3545 TGGAGCAGGCGCCTTCCTGCAACATCGGCGTCTGTCCGGTGGGGCAATTCAATTTTGAAC 3604
Db    21010 TGAGCGTTGCCCCCGCCCACAAGATGGGTTTGTGCCCGATCGGCGTGATGGACTTCGAGC 21069

Qy     3605 AGGTTCGGCCGGTTCTCGACCTGCGACATTCGGACGTTTACGTGCACGGCATGCTGGGCG 3664
Db    21070 CCCTGCGCTCGCAGCTCGCGTTGGACGAGCACCACGTGCTTCTGCACAGCTTCCTGGGGG 21129

Qy     3665 GGCGGGTAGACCCGCGGCAGTTCCAGGTCTGTACGCTCGGTCAGGATTCCTCACCGAGGC 3724
Db    21130 GCGGCGCAG---------------------------------------TGACGGCGGC 21148

Qy     3725 GCGCCACGACGCGCGGCGCCCCTCCCGGCCGCGAGCAGCACTTCGCCGATATGCTTCGCG 3784
Db    21149 GGCTGCGGCCAAGCCAGCACCCAAGGCGGCGGATGCTTCACTGC-CCGACGAGTTGAGGC 21207

Qy     3785 ACTTCTTGAGGACCAAACTACCCGAGTACATGGTGCCTACAGTCTTCGTGGAGCTCGATG 3844
Db    21208 GTTACCTGGCGTCGAAGCTGCCGGAGTACATGGTTCCGTCCTCGTTCGTTCTGTTGGATT 21267

Qy     3845 CGTTGCCGCTGACGTCCAACGGCAAGGTCGATCGTAAGGCCCTGCGCGAGCGGAAGGATA 3904
Db    21268 CCCTTCCGCTGACGTCCAACGGCAAGGTGGATCGAAGCGCGCTTCGTGCTCCGGCCGAAC 21327

Qy     3905 CCTCGTCGCCGCGGCATTCGGGGCACACGGCGCCACGGGACGCCTTGGAGGAGATCCTCG 3964
Db    21328 TGGTGGAAGTCAAGCCGCCCGCTGCGCGTGCGCCAGGCTCGGACGTGGAACGTTCCCTGG 21387

Qy     3965 TCGCGGTCGTACGGGAGGTGCTCGGGCTGGAGGTGGTCGGGCTCCAGCAGAGCTTCGTCG 4024
Db    21388 CCGCGATCGTTCAAGAGGTGCTCCAACTGCAAGAGGTGGATGTCCACCGCAACTTCTTCG 21447

Qy     4025 ATCTTGGTGCGACATCGATTCACATCGTTCGCATGAGGAGCCTGTTGCAGAAGAGGCTGG 4084
Db    21448 ATCTGGGGGGCACGTCGGTTCACATCGTGCAGATGCATCGCAGGATGAAGGAGCGGCTCA 21507

Qy     4085 ATAGGGAGATCGCCATCACCGAGTTGTTCCAGTACCCGAACCTCGGCTCGCTGGCG 4140
Db    21508 AGGCCGATATCCCGATCGCGCAGATGTTCCGGTACACCACCGTCAGTGCGCTGGCG 21563
```

RESULT 11
AF210249

```
LOCUS       AF210249                 77457 bp    DNA     linear   BCT 29-AUG-2000
DEFINITION  Streptomyces verticillus bleomycin biosynthetic gene cluster,
            complete cds.
ACCESSION   AF210249 AF149091
VERSION     AF210249.1  GI:9937210
KEYWORDS    .
SOURCE      Streptomyces verticillus
  ORGANISM  Streptomyces verticillus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 74421 to 74693)
  AUTHORS   Du,L. and Shen,B.
  TITLE     Identification and characterization of a type II peptidyl carrier
            protein from the bleomycin producer Streptomyces verticillus ATCC
            15003
  JOURNAL   Chem. Biol. 6 (8), 507-517 (1999)
  MEDLINE   99352421
   PUBMED   10421758
REFERENCE   2  (bases 1 to 77457)
  AUTHORS   Du,L., Chen,M., Sanchez,C. and Shen,B.
  TITLE     An oxidation domain in the BlmIII non-ribosomal peptide synthetase
            probably catalyzing thiazole formation in the biosynthesis of the
```

```
            anti-tumor drug bleomycin in Streptomyces verticillus ATCC15003
  JOURNAL   FEMS Microbiol. Lett. 189 (2), 171-175 (2000)
  MEDLINE   20389599
   PUBMED   10930733
REFERENCE   3  (bases 1 to 77457)
  AUTHORS   Du,L., Sanchez,C., Chen,M., Edwards,D.J. and Shen,B.
  TITLE     The biosynthetic gene cluster for the antitumor drug bleomycin from
            Streptomyces verticillus ATCC15003 supporting functional
            interactions between nonribosomal peptide synthetases and a
            polyketide synthase
  JOURNAL   Chem. Biol. 7 (8), 623-642 (2000)
  MEDLINE   20500448
   PUBMED   11048953
REFERENCE   4  (bases 1 to 77457)
  AUTHORS   Du,L., Sanchez,C., Chen,M., Edwards,D.J. and Shen,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1999) Chemistry Department, University of
            California at Davis, One Shields Avenue, Davis, CA 95616, USA
COMMENT     On Aug 29, 2000 this sequence version replaced gi:5326869.
FEATURES             Location/Qualifiers
     source          1..77457
                     /organism="Streptomyces verticillus"
                     /mol_type="genomic DNA"
                     /strain="ATCC15003"
                     /db_xref="ATCC:15003"
                     /db_xref="taxon:29309"
     CDS             complement(1199..2185)
                     /codon_start=1
                     /transl_table=11
                     /product="SyrP-like protein"
                     /protein_id="AAG02342.1"
                     /db_xref="GI:9937211"
                     /translation="MTGSVTLTPLGGIIPRPRGEGLTTGAEYDLGPLGDAGPDWVRAH
                     GPRLRERLATDGLILLHGLPTDGDGVDGFHDVVGSVGGDPLPYTERSTPRSVVKGNIY
                     TSTEYPADQPIPMHNENSYAAHWPSTLYFFCHTAPDTGGATPIADGRAVLDLIPAEVR
                     RRFSQGVVYTRTFRADMGLSWQEAFQTEDRGDVERHCRAHGQEFSWDGDVLRTRHHRP
                     ATAVDPGTGAEVWFNQAHLFHPSSLDPDLRQVLLETYGENGLPRDALFADGTPIPDAD
                     LATVRAAYTRAALALPWREGDIMLVDNLRMAHGREPFTGERRVLVAMTSADS"
     CDS             complement(2182..3918)
                     /codon_start=1
                     /transl_table=11
                     /product="peptide synthetase NRPS12"
                     /protein_id="AAG02343.1"
                     /db_xref="GI:9937212"
                     /translation="MTLTLRDAFLDQAARTPDAHAVVHGDTVWTYRELELRAGRMART
                     LAARGAGPGTLVAVRLPRGPEPVAALLAVVLTGAGYVPLADDDPPDRCRHILDDCAAA
                     LLLAEHPSRDGRTLTPDEALAPARPFDAAPVRAGDPAYVIYTSGSSGRPKGVLVEQGA
                     LGAYLAQARARYDGLSGRTVLHSSLSFDMAVTSLWGPLVSGGAIHVLDLKAIASGTQP
                     PPAASARPSFLKVTPSHLPLLGLLPDSCLPTGQLVIGGEALTGSALGPWRAAHPDVTV
                     VNEYGPTEATVGCCAYTVRPGDAVDPGAVPIGRPFAGTRLYVLDADGEPVAVGGVGEL
                     HIAGDQLARGYLGRPRLTEERFVPDPFAADGSRMYRTGDLVRERPDGDLEYLGRADGQ
                     VKVSGYRIEPGEIEAVLRGHAGVRDCAVVAVGEADARRLVAYVVPDPDSPPGTAAPAR
                     HAAEALPPYMVPATFVTVPELPLTPNGKLDRDALPGPPAGDAGPGDRTPAETLLCELL
                     ARALGIPEIDADADFLTSGGTSITALKLVAGARRVGIRLELTTVLRERTVRRILAAQP
                     DAASPLAEGVPE"
     CDS             complement(3915..4733)
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG02344.1"
                     /db_xref="GI:9937213"
                     /translation="MSKNAAHWSRIRTGDAPGVVLAVDFYGTGRQEATFRHLCDLLTD
                     PVEVWHAVPPAPDGDWSTATGAGHLRWWTEGLDTVLAGRPVRALVGYCAGGVFASALA
                     DALVEREGHRPRVVLFNPSAPGVATLTRDFRGLIAGMDLLTDGERAALLAETTAIRRA
                     HAPDALVPVAERYAALYREGCDLLCERLGVDASFGAELAAVLHSYLAYLTAALDVPPT
                     PLWRGAVSLTSREHQGTDFTDVEHGFDVARAELLSSPQVVAALTALLREHEASR"
     CDS             complement(4741..6024)
                     /codon_start=1
                     /transl_table=11
                     /product="putative transporter"
                     /protein_id="AAG02345.1"
                     /db_xref="GI:9937214"
                     /translation="MTARVDATPTYLAVLAVREARAPLLGSCLARMSFAVLPLALLLS
                     VRDATGSFAVAGLTSGALSATLTLFAPARARLIDRRGSRSGLVRLTVPYLLGLAVLIT
                     LAEAEAPTAALLVAAAVAGVFAPPLGPTMRVLWARILHGRQPLLHTAYALDSVTEEVV
                     FTVGPLLAGGLIAVAAPLASMITVMVLIAAGTACFVLSAATAAAPASGEADEDRPHGR
                     PMALPGMRTIVLSFGGVGLVVGVLQVVLPFIADHAGSPGAGGILLSMLSAGSAVGGLA
                     YGRIAWRSTPVRRFVVLVTGFTLAVLPLCLTASPVPAGAFALLVGLCLAPLFTTAYLL
                     VNDLVTASGTAPTEANTWVSTANNGGFAAGSAAAGVLLDSRGPTVTVTAAFAVAAATA
                     VMTVLRRRTLLLGAGHPEPAAATPADRTAPAEAEE"
     CDS             complement(6021..6773)
                     /codon_start=1
                     /transl_table=11
                     /product="thioesterase"
                     /protein_id="AAG02346.1"
                     /db_xref="GI:9937215"
                     /translation="MDFPLTRVNPWFSGGCDGRPRVRLCALPYAGGTAAVFKDWPAAL
                     PPGVELLTAHLPGRGDRFTEPPPATLEETAERLCEALPPSDLPTVVLGHSMGALLGYE
                     VAARLAARGRAPNLLIAAACRPPHVPPDASGPVTEAELAATLRAERPWDTALRDEELM
                     EAVLPALVADITAGDRYHRPRPRPLDLPLKVYIGADDDGTDWRTTLGWRACTARDCEV
                     VVLPGGHYFLETDRAAVLTRVATDLAEAEVGA"
     CDS             complement(6784..7797)
                     /codon_start=1
                     /transl_table=11
                     /product="clavaminic acid synthase-like protein"
                     /protein_id="AAG02347.1"
                     /db_xref="GI:9937216"
                     /translation="MTDLPLRTVALTGEESAEVDDLLRTLADVPVDSTVGLLHRTRLA
                     AQELPLRIRAELTGMRLYDSPRALVVTGFGVDDERIGPTPAARPAPDPERTRDLELLL
                     LLHAALLGEAFGWATQQNGRLVHDVLPVPGEETAQMGSSSETELLWHTEDAFHPLRCD
                     YVGLLCLRNHQRAATTVGWPDLSRLTTEDRAVLLEPRYLIRPDTSHTPAQNATGTRSA
                     ERFAAIAEMDDAPERVAVLFGDPEDPYLRIDPAYMSPAPGDAAARRAYDTVTALIEDE
                     LRHVVLDAGSLLLVDNYQAVHGRKPFAAAYDGRDRWLKRVNITRDLRRSRSARRSATS
                     LLV"
     CDS             complement(7859..9214)
                     /codon_start=1
                     /transl_table=11
                     /product="putative transporter"
                     /protein_id="AAG02348.1"
                     /db_xref="GI:9937217"
                     /translation="MMKSSRLRDRQLGGEDPVVAQESPQDAGPTPCQGDDGLNVFAAL
                     AALLEVEVPVRPLPHHAGLGRHVEHSSQAAAVALGPMQVASATTGVAGDGHQAGRRGQ
                     VTGVGVGREVAGGDDELGAEDRPHARQRLDDLGLWMAAERLADLLVNPLQTVVQGQDL
                     RGQVSHDLGGDVLPGQRGLLSLGSLQRRRGDGVGTAHASVGQPGRQPGPAAAAESCRG
                     LVARQQDQRALLRAVVEGPFQRGEDAGQRVAETVDHPDPVGHEVGTVGGQQREVGGQL
                     GGHVDRREVPSVAGGFGDDVGVAGVGLRLAPVSAGHAVDRAAGHVDGLLAVGREGQQQ
                     QRGGRAGDVHCPVDLVGQVEDLTHGGQDRRLIVADLLRPQRHTGLVDHRRPVMPLARV
                     DPGPDPGPSLAHSPLLTPNSIPSTRGTPRCHLRKKRPKRTSQSAARAPRRTGRPLLVS
                     H"
     CDS             complement(9211..17736)
                     /codon_start=1
                     /transl_table=11
                     /product="peptide synthetase NRPS11-10"
                     /protein_id="AAG02349.1"
                     /db_xref="GI:9937218"
                     /translation="MASDALSAHLEVSDCRQRGGITVTNDGSELAGQNVAAVRFERYS
                     AIAPERTAILHKGAATGYDELNRRAELTATRLADAGAGPSTLVAVALPRDPDLVATLC
                     ALLKLGAACLPLDPGIPAGRLREIMADASPDVLVTTRAVAPAFTGDGPVLFLDDAPPT
                     CSAVLPRHSAGTASEIAYVLYPTTPDEKSENSVVSYRDMARYLDDPTAGIPARAEILR
                     LVAPLLSGGRLVLDADETRPRPVTREAPRDMVEDVVAQVWCAVLGVDRVGVRDRFFDL
                     GGKSLAAVQVVARLRKLLGVELPLRALFDAPTVEELAARVRAEQAGGQGVREEAALEP
                     VGRSEPLPLSFAQQRLWFLDRLMPDRAFYTMCDAFRVRGGIDLGALRRALRMLVGRHE
                     TLRTAFVERDGVPYQLVGPADGPGARRVAAPTRVDLSLLEPAEREEAVRNLVAAEART
                     PFRPADGALLRVVVARLADDDHVLVVSTHHIVSDAWSVGVLVDELGRLYRECVTGDPA
                     ALPPPAVQYADFAVWQRAWMAGPVQEEHLAYWKRALDGAPSVLRLPMDHPRPAVQSER
                     GETVGFALPDALVAALEKLGREQGATLFMTLLGAFQVLLARHAGQEDIVVGVPAAGRT
                     RTETEPLVGFFVNTLPLRAICAPGLSFRDLLDQVREAALGAFAHQDLPFEALVEALAP
                     ERDLGHNPLVQVTFQLLGTPAARPDLIGTEVERYPVQEAVSQFDLSLDIKRADDGSYR
                     GILNYCPDLFDRRRMEVLVGHYLTLLGAAAADPGRPIGELPLSDGAERLRLLDGFGKR
                     DAAYAGPGSVPERFAEVARTAPDARAVTCGATTLTFAELNDRVERLAQALLGAGVTRE
                     TPVAVRLPRSTDSVVALLAVMRAGGVYVPLDPDWPADRTAYILDDTAASVVITRDLPA
                     LPGRLHVDPRRPAADGLVPAPRIDPDQAAYVIYTSGSTGAPKGVVVRHRSLNHLTSAL
                     QATFLGHDPYLAGADGVPPGDAKLRTTLTAPFTFDASMEQLSWMLAGHELFIVPEDVR
                     RDPSALVRFVREHRIDVIDTTSSQLELLVSHGLLDGEWAPSMVMVGGEAVSPSLWRTL
                     RDQRRTRCFNLYGPTEATVDATCHDLSDPADVPVIGTPLPHTHVRVLDDRLRPVPVGV
```

```
                    AGEIYLGGTGLARGYLNRPALTARRFVADPYPDTPGSRLYRTGDRARWRPDGTLEYLG
                    RTDDQIKIRGFRVEPGEIEAVLTHHPAVKEAAVVDDAHARLVAYVTLAEGGGAGPTDV
                    RRFAQGRLPAHMVPSAVVVLEALPLTSNGKLDRARLPAPAAGRPELDVRFVAPRDMVE
                    EVVAQVWCAVLGVDRVGVHDDFFELGGHSLLVVQVMTRIRKLLGVEVPLRELFDAATV
                    EELAARVRAARTEGLGRGAAPPLGPVDRSGPLPLSFAQQRLWYLDQLAPDSVSYNMCD
                    AYRVRGPLDLDALRRALRTLVERHETLRTAFVERDGVPHQVVSAPDAPAARRAAEVVR
                    IEAAGRTDEAVRDLVAAEARTPFRPADGALMRVAVARLADDDHVLVVTTHHIVSDGWS
                    VDILVDELGRLYREHVTGDPAGLPPLDVQYADFAVWQRSWMTGPVREEHLAYWKRALD
                    GAPSVLRLPADHPRPAVQSQRGETVEFPLPAPLVARLEALCREQGVTLFMALFGAFQV
                    LLARYSGQDDVVVGVPTANRTRAETEPLVGFFVNTLPVRVACSPELSFRALLDRVREA
                    ALGAFAHQDLPFEALVEALAPERDLGHHPLVQVTFQLLDAPDERLVLHGTDCVSLGFG
                    GVTSRFDLSLDVVSGRRGKRCVLTYCPDLFDRPRMEVLAGHYLTLLGAAADDPGLRVG
                    DLPLSDDVERLRLLGGSRPRYLPAPGAETVPDAFAAQVRATPDAPALVHGDSTLTFAE
                    LDTRVTALAVRLRRCGVAAETPVAVCLPRSADAVVALLAVLRAGGVYVPVDPEWPSGR
                    VAHVLDETAAPVVITRDLPADPGRVHLDPRQAPADDRDPLPRLHRDQAAYIIFTSGST
                    GAPKGVVVRHGSLYHLLGHVRRMAEGGPRRNVAHTTAMTFDPSLEQFLWLVAGHTLHV
                    APEEVRRDPEALVALVRRAAIDVLNVTPSHLTLLIEAGLLEGDRVPGTVLVGGEAVPA
                    ALWRTLRERTGATRFFNLYGPTEATVDATCHDLSDPADVPVIGTPLPHTHVRVLDDRL
                    RPVPVGVAGEIYLGGTGLARGYLNRPALTAQRFVADPYPDTPGSRLYRTGDRARWRPD
                    GTLEYLGRTDDQIKIRGFRVEPGEIEAVLTHHPAVKEAAVTVATDDGAARLVALVVPA
                    PRAPHGDSADGAPDAQVEEWNAVFEATHTDAADGELTFNIKGWNDSLTGAPIPAEHMR
                    EWVDTTVARLLERPAERVLEIGSGTGLLMWRLLPHVTEYTGTDFSRPAVDWLRDGLRR
                    RPAHRVRLLHREATDFTGVRAASTDLVVVNSVVQYFPDRAYLDTVLARALDATADRGR
                    VFVGDVRNLALAPQFYARQALAHAGPGAAARDVARAAGEFAAMDGELLVSPAYFAALA
                    ARSPRVTGVEILPRRGRHRNEMSLYRYDVVLHVGGDRPAAPEAEVLTWGDQVHDLASL
                    SARLGRGGPDALLVRGVANDRLTRDNELLDAPARTTAVEPEDLWGLADSTPYRVSVSW
                    AAADPRGAMDVLLVRRDAHDDGPLLVPHPVPEPSAPLTNTPTRHPSARQGGSAADGLR
                    SWLAERLPAHLLPARITEVDALPRTGTGKLDRGALGGLVTAGRGARAGDRPATAPRTG
                    LERTLADAWARVLGLPEVGVHENFFALGGDSLLAVRAVARCRRAGVRLTVRQLLSEQT
                    VAALAAALEEESQ"
     CDS            complement(18307. .18660)
                    /codon_start=1

  Query Match           17.2%;  Score 728;  DB 1;  Length 77457;
  Best Local Similarity 55.9%;  Pred. No. 3.8e-105;
  Matches 1545;  Conservative    0;  Mismatches 1185;  Indels   36;  Gaps    7;

Qy        182 TCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCCT 241
Db      61043 TCGTCACCGACCCCGGCGCCCGGCACGAGCCGTTCCCGCTCACCGACGTCCAGCGGGCCT 61102

Qy        242 ACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGG---ATCCACGCCTATCGCG 298
Db      61103 ACTACGTGGGACGCGAGGGCGGGTTCGCCCTCGGCGGCGTCTCCACCCACGCCTACCTGG 61162

Qy        299 AATACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCG 358
Db      61163 AGATCGAGGCCCCGCGGATCGACGTCGCACGGTTTACCGGCGCGCTGCGCGGGGTGATCG 61222

Qy        359 CGCGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTA 418
Db      61223 CCCGGCACCCCCATGCTGCGCGCCGTGATCCGTCCCGACGGGCTCCAGCAGGTGCTCACCG 61282

Qy        419 AAGTCGACGC---CGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGG 475
Db      61283 ACGTCCCCCCGTACGACGTGGCCGTGCACGACCTGCGCGACCTGGACGAGCCCGCGCGGC 61342

Qy        476 AAGCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCC 535
Db      61343 AGCGCCGACGCGCCGCGCTGCGCGAGGAGATGTCCCACCAGGTGGTGCCCGCCGACCTCT 61402

Qy        536 CTCCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCA 595
Db      61403 GGCCCCTGTT---CGACGTCCGCGTCTCCCTCGGCCCCACGGACGCCCTCGTCCACGTGG 61459

Qy        596 GTATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCA 655
Db      61460 GGGTGGACGCGCTGATCTGCGACGCCCACAGCTTCGGCCTCGTCCTGGCCGAACTCGCGG 61519

Qy        656 GCTTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATG 715
Db      61520 CCCGTTACGCCGACCCCGCACGCCGCTTCCCGCCCCTGACGGCGGACTTCCGGGACCACG 61579

Qy        716 TACTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGA 775
Db      61580 TCCTCCATCAGGAGGCGCTCCGCGGAACCGCCGAGTACGCGGCGGCGGAGCGGTACTGGC 61639

Qy        776 AGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTA 835
Db      61640 GCGAACGCCTGCCCGAGCTGCCGCCCGGCCCCGAACTGCCCCTGGCCGTCGCGCCCGAGA 61699

Qy        836 CCCTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTC 895
Db      61700 CCCTCGGCACCCCGCGCTTCACCCGCCGCTCCGGCCGGCTGGACGCGGCCTCCTGGACGG 61759

Qy        896 GATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCAT 955
Db      61760 CGGTCAAGGACCGGGCCCGCCGCGCCGGGCTCAGCCCCTCCGGCGTACTGCTGGCGGCGT 61819

Qy        956 TTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCT 1015
Db      61820 TCGCCGAGGTGATCACCGCGTGGAGCGGCCGGCCGCGCTACTCGCTGATGCTGACGGTCT 61879

Qy       1016 TCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGG 1075
Db      61880 TCGACCGCCCGCCGCTCCACCCGGACCTCGGGCGGATCGTCGGCGACTTCACCTCGCTCA 61939

Qy       1076 TCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTC 1135
Db      61940 GCCTGCTGGAGGTCGACCACAGTCGGCCGGCGACTTCACCGACAGGGCCCGCGCCCTCC 61999

Qy       1136 AAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGT-------- 1187
Db      62000 AGCGCCGCCTGTGGCAGGACCTCGACCACCTGGCGGTCGGCGGCGTGACGGTGACACGGG 62059

Qy       1188 -------CCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCG 1240
Db      62060 AACGGGCGCTGCGCCACGACGCCCGACCCGGTCTGCTCACACCCGTCGTCTTCACCTCCG 62119

Qy       1241 TGGTGCTCACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAGGCTCG 1300
Db      62120 ACCTGCCTGTCGGCGAGACCGCGGCCGAGGACGCGGACGGGGGAGAGGGATGGCGCTCG 62179

Qy       1301 GAACTCCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACG 1360
Db      62180 GAGAGCCCGTCTACGGCGTCAGCCAGACCCCGCAGGTCCATCTCGACCATCAAGTCGCCG 62239

Qy       1361 AGCACGATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACC 1420
Db      62240 AAGACCGAGGGGAGTTGGTCTTCAACTGGGACGCCGTGGAAGACCTGTTCGCCCCGGGCG 62299

Qy       1421 TTCTGGACGACATGCTCGAAGCGTACGTCGTTTTTCTCCGGCGGCTCACTGAG------G 1474
Db      62300 CCCTGGACGCCATGTTCGCCGCCTACACCGCCTCGCTGACCCGCCTGGCCCGGAGCCCCG 62359

Qy       1475 AACCATGGGGTGAACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGA 1534
Db      62360 AAGCCTGGCGGCGGCCCGGCACGCCGCCGCTGCCCACCGCCCAGGCGGCCGTGCGCCGGC 62419

Qy       1535 GCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGC 1594
Db      62420 GCACCGCCGCGACCGAGGCGCCCCTGCCCGCCCGCCTGCTGCACGAGGCCGTCGGCGACG 62479

Qy       1595 GGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACG 1654
Db      62480 CGGCCCGGCGCCACGCCGACCTGACCGCCCTGGTCGACGGCGACACCCGGATGACCTACC 62539

Qy       1655 AAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCC 1714
Db      62540 GGCGACTGACCGAGCACGCCCGGCGCGTCGGCCGCACGCTGCGCCGCCTCGGCGCCCGCC 62599

Qy       1715 CGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCG 1774
Db      62600 CCGGCCGCCTGGTCCCGGTGGTCGCCCGCAAGGGGTGGCGGCAGGCCGTCGCCGCGCTGG 62659

Qy       1775 CGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTA 1834
```

```
Db     62660 GCGTCCTGGAGTCGGGGGCGGCGTACCTGCCCCTGGACCCCGAACTGCCCGCCGAACGGC 62719

Qy      1835 TCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATG 1894

Db     62720 TCGTCCACCTCGTACGGCGCGCCGAAGCCGCCCTCCTCCTCACCGAACGCGCCCTGCTGG 62779

Qy      1895 GCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAG 1954

Db     62780 ACACGCTCGCCGTCCCCGTCGGCGTCAC---CGTGCTCGCGGTGGACGACGACGCGGCCC 62836

Qy      1955 GCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCA 2014

Db     62837 TCGACGCCGACGGCGGCCCGCTGCAGAGCGTGCAGAACCTCACCGACCTGGCGTACACCA 62896

Qy      2015 TCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCG 2074

Db     62897 TCTTCACCTCGGGCTCCACCGGCGAACCCAAGGGCGTCATGATCGACCACCTCGGCGCGG 62956

Qy      2075 TCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGG 2134

Db     62957 CCAACACCCTGGAATGCGTCAACCGCCGCTTCGGCACCGGCCCCGGCGACGCGGTCCTCG 63016

Qy      2135 CGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGG 2194

Db     63017 CCGTCTCCTCCCCGAGCTTCGACCTCGCCGTCTACGACCTGTTCGGCGTGCTGGCCGCCG 63076

Qy      2195 GCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGT 2254

Db     63077 GCGGCACCGTGGTCGTCCCCGCCCACGACCGCCGGCGCGACCCCGGACACTGGGCCGAGC 63136

Qy      2255 TGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCG 2314

Db     63137 TGATCCGGCGCGAGCGGGTCACCCTGTGGAACTCCGTCCCCGCGCTGGGCACCCTGCTCA 63196

Qy      2315 TCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGA 2374

Db     63197 CCGAGTACGCCGAGGCCCTCGCCCCCGACGCCCTGCGCACCCTGCGGGCGGTGCTCCTCA 63256

Qy      2375 GCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCAT---CAGGCCCGGCG 2431

Db     63257 GCGGCGACTGGATCCCCCTCGGACTGCCCGACCGGATCCGCGCCCTGTCCGCCCCCGGCG 63316

Qy      2432 TGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCG 2491

Db     63317 CCACCGTGATGAGCCTCGGCGGCGCGACCGAAGCCTCCATCTGGTCGGTCTGGTACGAGA 63376

Qy      2492 TGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGA 2551

Db     63377 TCGGGAAGGTGCACGAGGCGTGGAGCAGCATCCCCTACGGCACCCCCATGGCCAACCAGC 63436

Qy      2552 CGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCT 2611

Db     63437 GGCTGGAGGTCCTCGACGAGCAGCTGCGGCCCCGGCCCGACTGGGTGCCCGGCGAGCTGT 63496

Qy      2612 ACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGA 2671

Db     63497 ACATCGGCGGCACCGGCGTCGCCAAGGGCTACTGGCGCGACCCGGAACAGACCTCCCTGC 63556

Qy      2672 GCTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCT 2731

Db     63557 GCTTCCCCGTCCACCCGGGCAGCGGGCAACGCCTGTACCGCACCGGGGACTTCGCCCGCC 63616

Qy      2732 ACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCG 2791

Db     63617 ACCTCCCCGACGGCACGCTGGAATTCCTGGGCCGGCAGGACGACCAGGTGAAGATCGGCG 63676

Qy      2792 GATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCG 2851

Db     63677 GATTCCGGGTCGAACTGGGCGAGGTCGAGGCGGCCCTCGGCCGACTGCCCGACGTCGCCG 63736

Qy      2852 ACGCGGTGATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGG 2911

Db     63737 CCGGCGCGGTGATCGCCACCGGTGACCCGCGGGGCGACCGCCGCCTCGTCGGCTTCGCCG 63796

Qy      2912 TCCCGG 2917

Db     63797 TACCGG 63802
```

```
RESULT 12
AL646066/c
LOCUS       AL646066               189050 bp    DNA     linear   BCT 11-APR-2003
DEFINITION  Ralstonia solanacearum GMI1000 chromosome, complete sequence;
            segment 10/19.
ACCESSION   AL646066 AL646052
VERSION     AL646066.1  GI:17428714
KEYWORDS    .
SOURCE      Ralstonia solanacearum
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
            Burkholderiaceae; Ralstonia.
REFERENCE   1
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Nature 415 (6871), 497-502 (2002)
  MEDLINE   21681879
   PUBMED   11823852
REFERENCE   2  (bases 1 to 189050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1..189050
                     /organism="Ralstonia solanacearum"
                     /mol_type="genomic DNA"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
     gene            170..3325
                     /gene="RSc1698"
                     /note="synonym: RS02885"
     CDS             170..3325
                     /gene="RSc1698"
                     /function="elements of external origin; phage-related
                     functions and prophages"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE BACTERIOPHAGE PROTEIN"
                     /protein_id="CAD15400.1"
                     /db_xref="GI:17428715"
                     /db_xref="SPTREMBL:Q8XYR0"
                     /translation="MRNIIGYGGGKGGSSSSTPTEAPDSLHSVAYARVLDLVSEGEIA
                     GLVNGLQSIYLEGTPLANADGTLNFQNVAVDYRPGSQDQEAIPGFPSVESESAVGVEL
                     TATMPWVRAIANTQLSAVRVQLSVPALSRADTSSGNINGYRVEYAIDLAVDGGALQQV
```

```
                     LATAFDGKTTSKYTRTHRIELPRAKTGWTIRVRRITPNANSGTIADVTRIESIAEVID
                     AKLRYPNSALIGVRIDARQFSSVPTRSYHLRGRIIRVPSNYDPATRTYTGLWDGTFKV
                     AYSNNPAWVFYDIVLHKRCGLGDRVSADMVDKWALYQIGQYCDELVPDGRGGQEPRFT
                     CNCYLQQRNDAYPVLQDLASVFRGMAYWASSNVVAVADMPASASYLFHPGNVVDGRFT
                     YAGSARRARKTVALVSWNDPGDQYRAKVEHVEDGDGIARYSIQQTEVTAFGCTSQSQA
                     QRVGQWILLTSRLETETVSFKVGLDAAVVMPGAIIEVADPARAGRTNGGRVRSADGRT
                     VVVDRPVVVREGDTLLLNLPDGTAQRRTISRVDGQSLTVSADWSQPVQAEAVWSVESA
                     DLKTQLFRVVSVTESDGLTFDISALEYNQSKFSAIDNGTRIEARPVSVLPPSVQPAPT
                     DVTLATYSVVDQGIAITTMMVAWQQAASAVAYTVEWRRDNGEWVTAGRTGSQSLEVRN
                     IYAGTYVARVRAINALDVASLPAYSAETRLTGKTSPPPVVGTLIATAIVFGIRLDWAF
                     PTGPLDVERTELWWSKAPDRASATKLGDFAFPANTHTLMGLAAGATFYFWARLVDKSG
                     NVGAWYPSGNGVLGMSSDRASDILSYMKGQIGKTQLAGDLLSAVDSIDPPMAGSDEDF
                     AGDGTVYAGIVSTQSVLEEANRAIAQQVTTMQATVAQSSAAVQVAQQAVAEQGGRLAA
                     MYTIKTQIAANGRIYLAGIGFGVENDKGVIESQVLIAADRFAVLHPNGNNVFTPFVVQ
                     GGQVFMDSAFIADGTITSAKIGDSIQSSNFVAGQTGWRLSKSGLFENNGNGAGGRRVD
                     TSALTQIYDANGTLRVRLGVW"
     gene            3662..4318
                     /gene="RSc1699"
                     /note="synonym: RS02888"
     CDS             3662..4318
                     /gene="RSc1699"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAD15401.1"
                     /db_xref="GI:17428716"
                     /db_xref="SPTREMBL:Q8XYQ9"
                     /translation="MPAGLTVYGEHGFVQITEAYGNLALRQKSTVTPDGSGRGSITFS
                     ASRPFVCIQSANPVALLGSSNSGADWTVKFSAPAQAPFTVYVFDEPAAPSGQPGLQVF
                     KADGTLAFDSGLAYLKVSGVVTPPSGAPAYGQAWEASPLTAGNYAACLSYTRTGIYAV
                     PGTDTLFVADHVYTTANGAGLKLLQYTHRGAWNQGDGSGLKIDTANPPQIVLVDVSQL
                     "
     gene            4387..4878
                     /gene="RSc1700"
                     /note="synonym: RS02889"
     CDS             4387..4878
                     /gene="RSc1700"
                     /EC_number="3.2.1.-"
                     /function="elements of external origin; phage-related
                     functions and prophages"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE BACTERIOPHAGE ENDOLYSIN PROTEIN"
                     /protein_id="CAD15402.1"
                     /db_xref="GI:17428717"
                     /db_xref="GOA:Q8XYQ8"
                     /db_xref="SPTREMBL:Q8XYQ8"
                     /translation="MPYTDPSVLGGRNLAAFLDMVGFSEGTDNGRQPTRDRGYDVLVG
                     GGLFVSYADHPRILVELPRLGIKSTAAGRYQLLARYYDPYRRQLRLVDFGPAAQDAIA
                     VQQIRERGALADIQAGRLAAAIAKCKNIWASLPGAGYGQHEQKFEALRAHYLLCGGQE
                     VKG"
     gene            4878..5186
                     /gene="RSc1701"
                     /note="synonym: RS02890"
     CDS             4878..5186
                     /gene="RSc1701"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE LIPOPROTEIN TRANSMEMBRANE"
                     /protein_id="CAD15403.1"
                     /db_xref="GI:17428718"
                     /db_xref="SPTREMBL:Q8XYQ7"
                     /translation="MTESEMAVAAKVGGSAALGSVIALRFLPGTWWQRMLSFVSSLGI
                     GCLAGGAAVERFLLVPGSYTHMLAVASAAVFGLAIVNNAMQQIPEILTDLRRRFLAKE
                     "
     gene            5189..5464
                     /gene="RSc1702"
                     /note="synonym: RS02891"
     CDS             5189..5464
                     /gene="RSc1702"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE TRANSMEMBRANE PROTEIN"
                     /protein_id="CAD15404.1"
                     /db_xref="GI:17428719"
                     /db_xref="SPTREMBL:Q8XYQ6"
                     /translation="MLLTSINQVANAIIFVGSLWAVLTHQVPTRTGGALVLALVNFAA
                     LGNIVAHGTCHSWPEVALNVAVALCVVWAFWRLEVRRRFSALGKGRG"
     gene            5464..6009
                     /gene="RSc1703"
                     /note="synonym: RS02892"
     CDS             5464..6009
                     /gene="RSc1703"
                     /EC_number="3.4.-.-"
                     /function="elements of external origin; phage-related
                     functions and prophages"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE BACTERIOPHAGE-RELATED SIGNAL PEPTIDE
                     PROTEIN"
                     /protein_id="CAD15405.1"
                     /db_xref="GI:17428720"
                     /db_xref="GOA:Q8XYQ5"
                     /db_xref="SPTREMBL:Q8XYQ5"
                     /translation="MPGIKGTAIAAAIAMLLGALVGAAGAWTWQAHRYDKDIATLQRD
                     QAEKDRQQARADEAQLRQEREKRAGLERQLGELDARHYEELNNAKTENARLAAALAAG
                     TRQLSIRARCTAPADGVPSTTSGAGLGDGGAATAILYGEDAADIQALVADADAIAAAL
                     RFFQGRERAIQTRDTSVDARR"
     repeat_region   complement(6117..7304)
                     /note="predicted by Homology"
                     /insertion_seq="ISRso18 or RS05932"
     gene            complement(6273..7238)
                     /gene="tISRso18"
                     /note="synonyms: RS02893, RSc1704"
     CDS             complement(6273..7238)
                     /gene="tISRso18"
                     /function="elements of external origin; transposon-related

  Query Match             15.0%;  Score 634.6;  DB 1;  Length 189050;
  Best Local Similarity   54.1%;  Pred. No. 1.5e-90;
  Matches 1513;  Conservative    0;  Mismatches 1234;  Indels   48;  Gaps    9;

Qy        170 CCGCAGAATCCATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACA 229
              | || ||  |||||||||  || || | | |   |||    || || || || || ||  
Db     137989 CAGCCGAGCCCATCGTGCATGCGCCGGACCACGCGCATCTGCCCTTCCCGCTGACCGAGC 137930

Qy        230 TCCAAGAATCCTACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGAT---CC 286
```

```
Db  137929 TGCAGCACGCCTACTGGATCGGACGCCAGGGCCACTACGCGCTGGGCGGCGTGGCCTGCC 137870

Qy     287 ACGCCTATCGCGAATACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTC 346
Db  137869 ACGCCTACCTCGAGGCCGATGCCGCCGACGGCCTCGACCTCGGCCTGCTCGAACGCTGCT 137810

Qy     347 GGAAAG---TCGTCGCGCGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGC 403
Db  137809 GGAATGCGCTGGTCGCGCGCCACGGCGCGCTGCGCCTGGTGATCGACGAGAGCGGGCAGC 137750

Qy     404 AGGTGATCGAGCCTAAAGTCGACGCCGACATCGAGATCATCG---ATCTGCGCGGGCTCG 460
Db  137749 AGCGCATCCTGCCGCGGGTGCCGGCCTACCGCATCCGCGTGGCAAACCTGGGCGCGGCCA 137690

Qy     461 ACCGGAGCACACGGGAAGCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCT 520
Db  137689 CCCCGCAGGCCCTGGCCGCGCACTGCGACGACTGGCGCCAGGCCATGTCGCACCAGGTGC 137630

Qy     521 ATGACACCGAGCGCCCTCCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAA 580
Db  137629 TCGATGCCGCGCAATGGCCGCTGTTCGACGTGCGCGCCACGCACCTGCCGGGCGGGGCCA 137570

Qy     581 CCCGTCTCGTGCTCAGTATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCT 640
Db  137569 CGCGCCTGCACATCGGCATCGACATGCTGATCAACGACGCCACCAGCGGCCAGATCATCT 137510

Qy     641 TCAAGGA------------CTGGCTCAGCTTCTACGAAGATCCCGAGACCTCTCTCCCTG 688
Db  137509 GGGACGAGCTCGCCGCGCTGTACCGCGCCGGCGGCGACCTCGAACGGGCCGGCCTGGCGC 137450

Qy     689 TCCTGGAGCTCTCGTACCGCGATTATGTACTCGCGCTGGAGTCTCGCAAGAAGTC----- 743
Db  137449 CATTCGAGATCAGCTTCCGCGACTACGTGCTGGCGAAGTACGTGCACAGCGAAGCGCGCC 137390

Qy     744 -TGAGGCGCATCAACGATCGATGGATTACTGGAAGCGGCGCATCGCCGAGCTCCCACCTC 802
Db  137389 GCGCCGCGCGCGAATCGGCAAAGGCCTACTGGCTCGGCCAGCTCGAGACCCTGCCCCCCG 137330

Qy     803 CGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTGAAGGAGATCCGCTTCCGGCACA 862
Db  137329 CGCCGCAACTGCCCCTGCGCGCCGAGGCGCTGCACCGCGCCGCCCCGCGATTCTCGCGGC 137270

Qy     863 CGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGCGCG 922
Db  137269 GCCAGCATCGCCTGAGCGCGCCGCAGTGGCAAAGCCTGCGCGACCGGGCCGCGGCATCGG 137210

Qy     923 GGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGAGCG 982
Db  137209 GCTGCACGCCGGCTTCGCTGCTGATCGCGGTGTTTGCCGAAGTCCTGTCGGCCTGGAGCA 137150

Qy     983 CGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAACCGGCTCCCCGTCCATCCGCGCG 1042
Db  137149 CCGAGCCGCGGTTCACGCTCAACCTCACCACCTTCGACCGCTTGCCGTGGCACGCCGACG 137090

Qy    1043 TGAACGATATCACCGGGGACTTCACGTCGATGGTCCTCCTGGACATCGACACCACTCGCG 1102
Db  137089 TGCCGCGCCTGCTCGGCGATTTCACCGCGGTCACGCTGCTGCCGCTGGACTGCGCCGCGC 137030

Qy    1103 ACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGGATC 1162
Db  137029 CCCTGCCCTTCGGCCAGCGCGCCGCCGCCGTCAACGGGGCCGTGCTCGAGCACCTGCAGC 136970

Qy    1163 ACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCCGCCCGGGTCCTGGGGATCCAAC 1222
Db  136969 ACCGCGCCTTCAGCGCCGTGGACGTGCTGCGCGAATGGAACCGCGGCCGCGAACGGCAGG 136910

Qy    1223 GAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGC------TTAACCAGCAAGTCGTTG 1276
Db  136909 ATGCGGTGTCCATGCCGGTGGTGTTCACCAGCCAGCTCGGCATGAGCGACCCCACCAAGG 136850

Qy    1277 GTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGCACGCAGACTCCTCAGC 1336
Db  136849 GCGCCGCGCGCGCAAGCGTGCTGGGCACCGTCGGCTACGGCATCAGCCAGACCCCGCAGG 136790

Qy    1337 TGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTCCTCGCGTGGGACATCG 1396
Db  136789 TGTGGCTGGACCACCAGGCCTGCGAGCTCGACGGCGCGCTGATCTACAACTGGGATGCGG 136730

Qy    1397 TCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAAGCGTACGTCGTTTTTC 1456
Db  136729 TCGATGCGCTGTTCCAGCCCGGCGTGCTCGATGCCATGTTCGACGCCTACAACCGGATGC 136670

Qy    1457 TCCGGCGGCTCAC------TGAGGAACCATGGGGTGAACAGGTGCGCTGTTCGCTTCCGC 1510
Db  136669 TAGAACGCCTCGCCGCCGACGCCGATGCCTGGCTCGAGCCGCTGCCCGCCCTGCTGCCGC 136610

Qy    1511 CTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATA 1570
Db  136609 AAGCCCAGCGCGAGGTGCGCGCGCGCGTCAATGCCAGCACGGCGCCGCTGCCCGAGCGCT 136550

Qy    1571 CGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGT 1630
Db  136549 GCCTCGACCAGCTGTTCTTCGACCAGGCAGCGCGCCAGCCGCAGGCGCCGGCCCTGATCG 136490

Qy    1631 CGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGC 1690
Db  136489 CCAACGACCTCACCTGGACCTACGCACAACTGGCCGGCTGGTCCCGCCGCCTCGCCAACG 136430

Qy    1691 GGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCT 1750
Db  136429 GCCTCGCGCAGGCCGGCGTGCAGCGCGGTGACCGCGTGGCGGTGGTGATGCGCAAGGGCG 136370

Qy    1751 GGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCG 1810
Db  136369 CGGAGCAGGTCGCCGCCTGCCTGGGCATCCAGGCCGCGGGCGGCGCCTATGTGCCGGTCG 136310

Qy    1811 ATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCG 1870
Db  136309 ATGCGCAGACCCCGGCCGCGCGCCTGCACGCCATCCTGCAAGGCAGCCGCATCCGCGTCG 136250

Qy    1871 TGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGC 1930
Db  136249 CGCTGACGCAGGCCGACTGCCTGCCGATCGTGCGCGAGCTGACGCAGGCGCTCGACGTGC 136190

Qy    1931 TCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGC------CTCCGATGATGCCCA 1984
Db  136189 GCGTCATCGATGCCTCGCACGACGCGGTCGCCGGCTGGCCGGACGCCCCGCCCACCGTCG 136130

Qy    1985 TTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCA 2044
Db  136129 CGCGCGATCCGCTCGACCCCGCCTACGTGATCTACACCTCCGGCTCGACCGGCACGCCCA 136070

Qy    2045 AGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCT 2104
Db  136069 AAGGCGTGCTGATCGACCACCGCGGCGCGGTCAACACGGTGCTCGACATCAACCGCCGCT 136010

Qy    2105 TCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGG 2164
Db  136009 TCGGCATCACCGCGGGCGATCGTGTGCTTGGCCTGTCGGCGCTGTACTTCGACCTGTCGG 135950

Qy    2165 TCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCA 2224
Db  135949 TCTACGATCTCTTCGGCAGCTTCGCGGCGGGCGGCACGCTGGTGCTGCCGCAGCCCGGCG 135890

Qy    2225 AGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGA 2284
Db  135889 GCGTGCGCGATCCGGCGCACTGGCTGGCCTTGGCCACGCGCCATGGCGTGACCGTGTGGA 135830

Qy    2285 ACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGC 2344
Db  135829 ACTCGGTGCCGGCGCTGCTCGAACTGCTGCTCGACGAGGCCGAAGCCGCCGGCCAGACGC 135770

Qy    2345 TCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTG 2404
Db  135769 TGCCCGGCATCGGCCACGTCT---TCCTGAGCGGCGACTGGATCCCCCTGCACCTGCCGG 135713
```

```
Qy       2405 GCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAG 2464
                |   || |  ||| ||  ||| | || |      |||||   | ||||||| |||||||||
Db     135712 CCCGCCTGCGCGCCCTCGCGCCGGCCGCGCGCCTGATCGCCATGGGCGGCGCCACCGAAG 135653

Qy       2465 CGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGCATCC 2524
              |||||||||||| | ||| ||  ||        | ||           |||     |||||
Db     135652 CGTCGATCTGGTCGAACTGGTTCCGGGTCGATGAGGTGCCGGCCCACTGGCACTCCATCC 135593

Qy       2525 CCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGC 2584
              |||||||    |||||  ||||||||     | || |||||||||        ||
Db     135592 CCTACGGCTACCCGCTCGGCAACCAGCGCTACCGCGTGCTCGACACACACCTGCGCGACT 135533

Qy       2585 GCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACT 2644
              |||| |  |    ||   | ||   | |||| ||| |||||    |||| ||||| || ||||||
Db     135532 GCCCCGACCACGTCACCGGCGATCTCTGCATCGGCGGCATCGGCCTGGCGCTCGGCTACG 135473

Qy       2645 GGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAGCGCC 2704
                |  ||||    |||    |||   |  ||||||| ||    || ||         ||| |||||||
Db     135472 AGAACGATCCCGAGCGCACGGCCGCGAGCTTCTTCCACCATCCGCTCGACGGCGAGCGCC 135413

Qy       2705 TCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGC 2764
              | |||     ||||||||| |||||||     |||||| || |      |||||||| | ||||
Db     135412 TGTACCGCACCGGCGACCTGGCGCGCTACTGGCCCGACGGCACGCTCGAGTTCCTCGGGC 135353

Qy       2765 GTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAA 2824
                |     ||| |||  ||||| |     |||   ||||    | ||||| || ||||||||
Db     135352 GGCGCGACTTCCAGGTCAAGATCGCCGGCAACCGGATCGAGCTGGGCGAAATCGAAAGCG 135293

Qy       2825 CGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGACGCGG 2884
              | ||   |   ||  |      | | ||||| || ||| | |     || ||||     |     |
Db     135292 CCCTGCTGCGCCACGCAGGCCTGCGCGATGCCGTGGTCGACGCCATCGGCCCGGCGCGCG 135233

Qy       2885 CGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAG 2919
                 ||||||| |||      || |  ||||| ||| ||
Db     135232 GCAACAAGCGCCTGGCCGCATGGGTGGTGCCGAAG 135198


RESULT 13
AF516145
LOCUS       AF516145               40156 bp    DNA     linear   BCT 03-OCT-2002
DEFINITION  Lyngbya majuscula barbamide biosynthesis gene cluster, complete
            sequence.
ACCESSION   AF516145
VERSION     AF516145.1  GI:23452290
KEYWORDS    .
SOURCE      Lyngbya majuscula
  ORGANISM  Lyngbya majuscula
            Bacteria; Cyanobacteria; Oscillatoriales; Lyngbya.
REFERENCE   1  (bases 1 to 40156)
  AUTHORS   Chang,Z., Flatt,P., Gerwick,W.H., Nguyen,V.A., Willis,C.L. and
            Sherman,D.H.
  TITLE     The barbamide biosynthetic gene cluster: a novel marine
            cyanobacterial system of mixed polyketide synthase
            (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an
            unusual trichloroleucyl starter unit
  JOURNAL   Gene 296 (1-2), 235-247 (2002)
REFERENCE   2  (bases 1 to 40156)
  AUTHORS   Chang,Z., Sherman,D.H. and Gerwick,W.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-2002) Microbiology, University of Minnesota, 421
            Delaware Street, SE, Minneapolis, MN 55455, USA
FEATURES             Location/Qualifiers
     source          1..40156
                     /organism="Lyngbya majuscula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:158786"
     gene            5531..5818
                     /gene="barA"
     CDS             5531..5818
                     /gene="barA"
                     /function="probably involved in the activation of leucine
                     to iniciate barbamide biosynthesis"
                     /note="peptidyl carrier protein"
                     /codon_start=1
                     /transl_table=11
                     /product="BarA"
                     /protein_id="AAN32974.1"
                     /db_xref="GI:23452291"
                     /translation="MSNSTELPISDVVVPQTETEKQLAEIWKDVLSVETISIEDRFMD
                     IGGNSINLVEVVNQVTEKMGVSIKARWFFDKHKSTIAELSKEIDAVREQTH"
     gene            5850..6752
                     /gene="barB1"
     CDS             5850..6752
                     /gene="barB1"
                     /function="probably related to chlorination"
                     /note="similar to the syringomycin SyrB2 protein and the
                     coronamic acid CmaB protein"
                     /codon_start=1
                     /transl_table=11
                     /product="BarB1"
                     /protein_id="AAN32975.1"
                     /db_xref="GI:23452292"
                     /translation="MKTRAKAELSYYLSVEEKQFFRDNGYIGPFTLFQPEEILELWNE
                     IRMDLLDRETAPFPNSKMNYDRHLDIQALSDIISHPKIAHRVSSLLGSNVLCWRTAWF
                     PKYPGDEGTDWHQAESFVEFEGKSKLVPTESHDGPWELTAWVAFSEATRENGCMKVMP
                     GTHNTWYFDEHRNIEFEPDKINQKLTGGKKTGVYGYDYYKLKLDPNWEPDESQAVHLE
                     MEPGQFILFTSRCMHGSEPNTSGSSIRYGWSTRFVPTDVRVYPDWESFQHFGEVFPLE
                     RYATVLVAGEDTYKHNKIRRPLGQ"
     gene            6812..7705
                     /gene="barB2"
     CDS             6812..7705
                     /gene="barB2"
                     /function="probably related to chlorination"
                     /note="similar to BarB1, syringomycin SyrB2 protein and
                     the coronamic acid CmaB protein"
                     /codon_start=1
                     /transl_table=11
                     /product="BarB2"
                     /protein_id="AAN32976.1"
                     /db_xref="GI:23452293"
                     /translation="MKTITAEPRYYLSVEEKQFFQENGYIGPFTLFPPEEMLELWYGI
                     KMDLLDKETAPFPNNKMNYDRHLDIKALNDIISHPKVVDRVSSILGPNVLCWRTEWFP
                     KYPGDEGTDWHQAESFVEFEGESKLVPTESHDGPWELTAWVTFSEATRKNGCMKVMPG
                     THTTWYFDEHRNIKFEPDNINQKLTKGQKSGFYGYDYDKLKLDPNWEPDESQAVHLEM
                     EPGQFFLFTSRCMHGSEPNTSESSIRYGWSTRFVSTDVRVYPDWDSFEHFGEKFMLDR
                     YSTVLVAGEDTYKHNKVRQPL"
     gene            7859..8665
                     /gene="barC"
     CDS             7859..8665
                     /gene="barC"
                     /function="proofreading or editing role or chlorination
                     role"
                     /note="thioesterase II"
                     /codon_start=1
                     /transl_table=11
                     /product="BarC"
                     /protein_id="AAN32977.1"
                     /db_xref="GI:23452294"
                     /translation="MQTDDSSWLIFDETPKCEASKMRLFCFHYAGAGGSIFQDWSRAL
                     PDEVELVAIQLPGREKRCYEPLMSNIEDINLVLAQELSPFLSKPFAFMGHSMGALIAF
                     DLIRLLRQQELGQPQFLMVSGFDAPQLRPKESSNMHQLPDDEFAQAIRDYNGIPQVIA
                     QNEELFELLLPRLRSDIKLGETYIYKDEEPLDCRIIALYSTEDSFVSKSGMYSWQEQT
                     KGEFELYTFPGDHFFLHSAEELVLNTICTELQRYIKANDNCSTKVALQKS"
     gene            8700..10253
                     /gene="barD"
     CDS             8700..10253
                     /gene="barD"
                     /function="adenylation; involved in activation of leucine"
                     /note="sequence similar to adenylation domains of NRPS"
                     /codon_start=1
                     /transl_table=11
                     /product="BarD"
                     /protein_id="AAN32978.1"
                     /db_xref="GI:23452295"
                     /translation="MGKNIANLVTRHSKKSPHKCAITDGQTELSYGVWEAASNRFARK
```

```
                     LIEMGCQKGDRVVVLASKQTILMVGILGIFKAGCIHVPLDGKMPVNRLNKIITEIIPS
                     VILTDQELLHLVSENVPQEVKIVLIEQLHPLITDSTVEEKDRLPQIESEDIAYCIYTS
                     GSTGDPKGVLIKHGSVIDFFEGTREFYDVDEDSRCVSFSPFHFDASVIDILFPLYQGA
                     WLYLYSDVVLHELLFEVIKNNSVTHPAAFGSMLGLIAQTKEFDDVILKDLKTILTGAD
                     VPDIKVIQKWLGKGENIKIINGYGPTEGTCACAAYIIRDRDPNRRELYPIGKPLKNAK
                     LYLVDQQNQEIYDPEIPGELLIGGTQVMAGYWNREQQSKERLVEFKGTTCYKTGDICK
                     YLADGNLFFVGRNDNEVNVGGYRVHLNGIKQVINRFHWVENSEVVTLESKYGEKMLAA
                     ALMLANETSKDSALKQIKEHLNQELPQYMVPRQIVVMEKFPQLSSGKTDRKTLSSILE
                     RDLLASKVH"
gene                 10313. .15433
                     /gene="barE"
CDS                  10313. .15433
                     /gene="barE"
                     /function="involved in activation of trichloleucine and
                     condensation with malonyl group"
                     /note="NRPS/PKS; A, PCP, KS and AT domains are located on
                     BarE"
                     /codon_start=1
                     /transl_table=11
                     /product="BarE"
                     /protein_id="AAN32979.1"
                     /db_xref="GI:23452296"
                     /translation="MKDNPMGIHQKNYDEVNHKLSTSFGGSLNIPSNSPTLLCDLLPR
                     AALTNKGITYIQADGSEQYQSYQDLLLDAKRILAGLRQWDLKPQDKVILQLEEGHDFI
                     PAFWGCLLGGFIPVPLSIAPTYQQLNNVTKKLHNAWQMLEHPLILTNRELAPEIRSLS
                     THLNLRDFQTGVIEDLRKQDPTQEIYRGQPDELALLLLTSGSTGLPKGVMLSHRNILS
                     ESAGTIQMNHFDSEAVTLNWMPLDHVGGIIMLNVMAVELGCQQVHAPTKYILQNPLHW
                     LALIERHQATISWGPNFAYSLINDLVQDAEAGSWDLSSISFLVNGGEQISPKVARRFL
                     KLMKRQGLAETVLRPAFGMSETSGGITWSDQFSLSNTSDDMPFVGLGLPIPGACLRIV
                     NDDNEIVHEGAIGKLQVKGASVTSGYYQNPEINQAAFTADGWFRTGDLGYLENGYLVL
                     TGRDKDVILINGVNYPAHEIEAVVEEVDGVEISFTAACAIQGKTNSSSSDQLAIFFSA
                     VEANIETNPEVLKGLIKKIRGAVVQNIGINPDYLLPVGKESIPKTAIGKIQRQKLKQQ
                     FEAGEFDPLIETVGQSLRIMGSQETALALPQTEVEQVIVNIWQDVLGLDKIGLQENFF
                     ELGGHSLLLLTLQSQLQESFGKELSIAEMFSHPTIEALAQYFTQTQPEDSAAQQGYAR
                     AQIRQEVRKADSTEIAVIGMAGRFPGADNLDQFWQNLQNGTESITFFSDSELLASGID
                     PKLLSNPNYVRASPILSDVEYFDAHFFGYTNREAEFMDPQQRLLMECAWESLEDAGYD
                     PLTYSGSIGLYAGAMLSTYFIHNLLPNRDTFDEHDDLRVLTPDSMGGLQVVLANDKDY
                     LTTRISHKLNLTGPSVNIQTACSTSLVTVHTAVQSLLNGECDLALAGGVSVQVPQRVG
                     HLHQEGLIASPDGHCRAFDAQAEGTIFGNGVGLVALKRLEEALQDGDHIYAVIKGSAI
                     NNDGSNKVAYTAPNGEGQTTVVAEAIATANINPETITYVEAHGTGTLLGDPIELGALS
                     QAFQAKTQRKQFCALGSVKTNVGHLQITSGIVGLIKTVLALYHKQIPPSLHFQTPNPK
                     IDFANSPFFVNTMLSEWERNGHPRRAGVNSLGFGGTNAHVILEEAPENAQPKTRNEDG
                     AKLEDRSHHLLTLSAKTAPALQALVKRYQRYLSDNTEVELGDICYTACVGRAHFEHRL
                     AIVSDSIPDLRCKLANVLDDNAELSPEQTQELAQGQTKASVRAQIAFLFTGQGSQYLN
                     MGRQLYQTQSVFRDAIDQCDQILQSELEVSLIDLLYPQTADEQNAARLNNTAYTQPAL
                     FAIEYALAQLWQSWGIQPAVVMGHSVGEYVAATIAGVFDLETGLKLIAARGRLMQQLP
                     AGXAMVSVMASEERVRSLISPHTEKISIAAINGPESVVISGVAEVLGAISNQLESAGV
                     KTKSLQVSHAFHSPLMEPILAEFTTVAEQLTYHSPRIPLISNVTGVKADEQIATAQYW
                     VNHIRQPVRFAHSMEALSEFGYDCFLEIGPKPILLGMGRQCLPEGTGVWLPSLRPTVD
                     DEQQMLSSLGKLYVQGVRVNWSSLNQEYSYQKVTLPTYPFERKKYWIEKYETNEQKFT
                     SATSKNQKNNFIRLLNQGNVLDLIQYLKKTLDLSREEEVKLLSKFAQLIINLYHQENE
                     LLTNKDRKEKDLMNQMLQMDSENLNKLLSEAEKFSQNS"
gene                 15476. .16990
                     /gene="barF"
CDS                  15476. .16990
                     /gene="barF"
                     /function="involved in addition of O methyl group"
                     /note="contains methyltransferase (MT) and ACP domains"
                     /codon_start=1
                     /transl_table=11
                     /product="BarF"
                     /protein_id="AAN32980.1"
                     /db_xref="GI:23452297"
                     /translation="MSDDLENMNNLQERLSKLSPEQRQFFEKKIAENKPHQQLSGQEF
                     IAEKGSTEERLSRYYRALDHSNTAKDYARFAPFSEVVPDFSWLEVILNPDQHPEQTSL
                     SLQAQAEMKAVLFRGIDFSSIKKVMDIGCGYSHDLIDLATNHVHLQLDGYNISPEQVK
                     AGEQKIQGLGYSDRIYLYNRDSAKQPLPDTYDLIFSCQVIHHIKRKEDVFLNISQHLN
                     DSGFFVAAEIISNLPLTPIDDAKSTAYYVTRSKWAQLLARNNLKVVEGVDASLGIANY
                     LYDPNFTDNLARLCQGYDENAKDHFVGPHELGELLRKKLAVYLLITVQKDSFTEKETI
                     LRLNQEKLQNLTPYSKIIENSPNGKPLLLPQFFREGSTTFSQGTREDFRQQFLSQDLS
                     DAQVGLAFYLKTVVAAILKETQAEINEETSLTVMGVDSLMVLELRNRVQADLGVDISV
                     ADLMGDINISRLAIKVSELLHRNAQGAQDLESKMDGDNLQEIALQDKKMREGRL"
gene                 16987. .25650
                     /gene="barG"
CDS                  16987. .25650
                     /gene="barG"
                     /function="responsible for the recognition and activation
                     of phenylalanine and cysteine and for the release of the
                     barbamide intermediate from the assembly line"
                     /note="NRPS; two NRPS modules are located on the protein
                     containing C, A, MT, PCP, Cy, A, PCP and Te domains"
                     /codon_start=1
                     /transl_table=11
                     /product="BarG"
                     /protein_id="AAN32981.1"
                     /db_xref="GI:23452298"
                     /translation="MNLSEFLQELVISGWQFWAEEGQVCFQAPADSTDQVLAQLKQH
                     KRDILTILQEHPEVLQVYPLGYGQQGIWFLWQLFPDNPNYNVSFATRIYSQVNVTTWQ
                     QTFEALRKRHPLLCSTFPKCGETPIRQHSEQLDFVQIDASTWDENELQTQVVAAHRHP
                     FDLQTDPVMRVRWFTRSEQEHILLLTIHHIAWDGSSANIIVKELSELYQAHCAGVAVD
                     LPSLQHTYQDYVKWQQQLVEGSKGESLWTYWQQQLAGELPVLNLPTDRPHPPIQTNNG
                     AVYRFQLPEHLVTQVKALSQAEGATLYMTLLAAFQVLLHRYTGQEDILVGSPTSGRTR
                     PEFTSVVGYFVDSMVMRAKVSGSLSFREFLTQVRQTVIDALAHQDYPFSLLVEKLQPE
                     RDLSRSPIFQVFFGLHNFLQSETQQLFLGETKTLVHWGGMEVETFLFDQYESLEDLVL

  Query Match             14.4%;  Score 609.4;  DB 1;  Length 40156;
  Best Local Similarity   52.9%;  Pred. No. 2e-86;
  Matches 1499;  Conservative    0;  Mismatches 1291;  Indels   45;  Gaps    7;

Qy        179 CCATCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAAT 238
Db      21581 CAATTGTGCCTGCGCCAAGCCAGCGCTATCAACCCTTTCCCCTCACCGAGATTCAACAAG 21640

Qy        239 CCTACTGGCTGGGCCGGACAGGAGCGTTTACGGTCCCCAGCGGGATC---CACGCCTATC 295
Db      21641 CTTATTGGCTCGGCCGCAATTCGCACTTTGACTTGGGTAATATTACCACGCATGGTTATC 21700

Qy        296 GCGAATACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCG 355
Db      21701 TGGAACTGGACTGTGAGAATTTAGCTTTAGACCGGTTAAGTCAGGCTTGGCAGCAAGTAA 21760

Qy        356 TCGCGCGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGC 415
Db      21761 TAGACCATCATGACATGCTGCGAATGGTGATTCTACCCAATGGTGAACAGCAAGTCCTTG 21820

Qy        416 CTAAAGTCGACGCC---GACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACAC 472
Db      21821 AACAGGTTTACCCCTATCAGATTGAAGTTTTAGACTTGCGCGGTCAACCAGAACAAATCG 21880

Qy        473 GGGAAGCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGC 532
Db      21881 TCTCAACCGAACTAGAAACCATTCGTTATCGCCTGTCCCACGAGATGTTCCCCGCTGGAG 21940

Qy        533 GCCCTCCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGC 592
Db      21941 AGTGGCCACTATTTAAGATTCGTGTGACTCGCCTAGCGGATCAGCGTTACCGCTTACACT 22000

Qy        593 TCAGTATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGC 652
Db      22001 GGAGCTTTGATGCCCTAATTGCTGATGCCTGGAGCATGATCATTGTTGGCAACAATGGC 22060

Qy        653 TCAGCTTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATT 712
Db      22061 TGCAGCTATACCAGAACCCAGACAGTTTCCTACCAAAATTAGACCTCACATTTCGGGACT 22120

Qy        713 ATGTACTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACT 772
Db      22121 ATGTGTTAGCAGAACTGAGTCTCAAAGACACGCCCCAATATCGACGCTCTCAACAGTATT 22180

Qy        773 GGAAGCGGCGCATCGCCGAGCTCCCACCTCCGCCGACGCTTCCGATGAAGGCCGATCCAT 832
Db      22181 GGTGGAATCGCTTAGAGACACTCCCCCTGCACCAGAACTTCCCTTGGTCAAGCAAACTG 22240

Qy        833 CTACCCTGAAGGAGATCCGCTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGG 892
Db      22241 CCACACTAGAGCAACCTGAATTTAATTGTTATCGGGCTGAACTGTCTGCCCCAGATTGGC 22300
```

```
Qy       893 GTCGATTGAAGCGGCGTGTCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTG 952
Db     22301 AGCAATTGCAAGCCAGGGCAAAACAAGCCAGTTTGACACCGTCAGGGGTCTTGTTAGCCG 22360

Qy       953 CATTTTCCGAGGTGATCGGGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGC 1012
Db     22361 CGTTTGCCGATTTCCTAGCCTATTGGAGCAAAAGTCCCAAATTCACGATCAATCTCACCC 22420

Qy      1013 TCTTCAACCGGCTCCCCGTCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGA 1072
Db     22421 TCTTTAACCGGCTACCTTTGCATCCCCAAGTCAATGATTTAGTGGAGACTTCACTTCCC 22480

Qy      1073 TGGTCCTCCTGGACATCGACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTA 1132
Db     22481 TGACACTCTTAGAGGTCAATCAAAAAAACGCAGCTCCCTTTGCTCAGCGTGCTCAACGTC 22540

Qy      1133 TTCAAGAGCAGCTGTGGGAAGCGATGGATCACTGCGACGTAAGCGGTATCGAGGTCCAGC 1192
Db     22541 TCCAGGGACAACTTTGGCAAGACCTAGACCACCGCTATGTTGGGGGCGTTGAGGTGCAAC 22600

Qy      1193 GAGA----GGCCGCCCGGGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTC 1248
Db     22601 GAGAACTTCGCCGTCAGCGCGGTAGTTATCAACCCATGGGGGTTGTATTTACCAGTACGC 22660

Qy      1249 ACGAGCGCGCTTAACCAGCAAGTCGTTGGTGTCACCTCGTTGCAGAG------------- 1295
Db     22661 TGGCGCTGAATACGTCGGCAGAGAAGGGATTGCCGTCCAATGAATGGCACGCTTGGCCAT 22720

Qy      1296 -------GCTCGGAACTCCGGTGTACACCAGCACGCAGACTCCTCAGCTGCTGCTGGATC 1348
Db     22721 TCGACCAATTGGGAGAAACTGTTTATATGGTCTCTAAAACCCCTCAAGTGTGGCTAGACA 22780

Qy      1349 ATCAGCTCTACGAGCACGATGGGGACCTCGTCCTCGCGTGGGACATCGTCGACGGAGTGT 1408
Db     22781 ATTCTGTTGCTGAACAAAATGGTGCCTTACTCCTAATCTGGAATGTAGTCGAAGACCTCT 22840

Qy      1409 TCCCGCCCGACCTTCTGGACGACATGCTCGAAGCGTACGTCGTTTTCTCCGGCGGCT-- 1466
Db     22841 TCCCAGAGGGGTTCCTCAATGACATGTTCACAAGCTACTACCATTGGTTGCAACAGCTCG 22900

Qy      1467 ----CACTGAGGAACCATGGGGTGAACAGGTGCGCTGTTCGCTTCCGCCTGCCCAGCTAG 1522
Db     22901 CAACCTCCGATGTGGCTTGGGCGCAGACTTGCCCTCAACTACTGCCCCTCTCTCAACTTA 22960

Qy      1523 AAGCGCGGGCGAGCGCAAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGCACGGCC 1582
Db     22961 CCCAACGCCTCCAAGTTAATGAAACCTATGCCCCTGTTTCTGAAGAGACCCTCCATAATT 23020

Qy      1583 TGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGCAGCTCGCCGTGGTGTCGGCGCGCAAGA 1642
Db     23021 TGTTTGTGAAGCAAGTTCAGCAGCGACCAGAGGCGATCGCCCTCATCACCCCGCAACGCA 23080

Qy      1643 CGCTCACGTACGAAGAGCTTTCGCGCCGTTCGCGGCGACTTGGCGCGCGGCTGCGCGAGC 1702
Db     23081 CCCTCACCTACCATGAACTCTACACCGAAGCTCAGGCGCTCGGCCAGCAAGTTCAGCAAT 23140

Qy      1703 AGGGGGCACGCCCGAACACATTGGTCGCGGTGGTGATGGAGAAAGGCTGGGAGCAGGTTG 1762
Db     23141 TGGGAGCCACCCCCAACACCCTAGTAGCCGTGCTCATGGAAAAAGGCTGGGAGCAAATCG 23200

Qy      1763 TCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCCTACGTGCCGATCGATGCCGACCTAC 1822
Db     23201 TTGCTGTCTTAGGCATTTTGATGGCTGGTGCAGCCTATCTCCCCATTGATGCCGCCCTGC 23260

Qy      1823 CGGCGGAGCGTATCCACTACCTCCTCGATCATGGTGAGGTAAAGCTCGTGCTGACGCAGC 1882
Db     23261 CCCAGGAGCGTCAATGGTCTTTGCTGGAGCAAGGAGAAGTCAAACTGGTCGTGACACAAG 23320

Qy      1883 CATGGCTGGATGGCAAACTGTCATGGCCGCCGGGGATCCAGCGGCTGCTCGTGAGCGAGG 1942
Db     23321 CTGCCCTAAACGCTAGCTTAGGATTGCCAGACCACCTGCACTGCCTGGTGGT------TG 23374

Qy      1943 CCGGCGTCGAAGGCGACGGCGACCAGCCTCCGATGATGCCCATTCAGACACCTTCGGATC 2002
Db     23375 CTTCCCAACCACAGGAGATCATAGACACTCCCCTGGAGGCAAACGTTAGCAGCAGCGATT 23434

Qy      2003 TCGCGTATGTCATCTACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATC 2062
Db     23435 TAGCCTATGTGATCTTTACCTCCGGTTCTACAGGAACACCCAAAGGGGTGATGATCGATC 23494

Qy      2063 ATCGGGGTGCCGTCAACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAG 2122
Db     23495 ACCGAGGAGCCGTTAACACCATCCAAGATATTAACCAACGGTTTGACGTTCAGCCCACTG 23554

Qy      2123 ACAGGGTGCTGGCGCTCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGA 2182
Db     23555 ACCGAATGCTGGCGGTCTCTGCCCTCAACTTTGACCTGTCTGTCTACGATATCTTTGGCC 23614

Qy      2183 TCCTGGCGGCGGGCGGTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGC 2242
Db     23615 TCCTGGCAGCGGGGGGCACCCTAGTCATGCCCACTCCAGAGGCAGCCAAAGACCCAGTCC 23674

Qy      2243 ATTGGGCAGAGTTGATCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGA 2302
Db     23675 ATTGGGTTGAGTTAATGACGACCCATCAAGTCACCCTCTGGAACACTGTCCCAGCCCTGA 23734

Qy      2303 TGCGGATGCTCGTCGAGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGC 2362
Db     23735 TGCAGATGCTAGTGGAATATCTGTCTGAGCACCCAGACCAAGTAACTGAGGACCTCCGAC 23794

Qy      2363 TTTCGCTGCTGAGCGGCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCA 2422
Db     23795 TTGCCCTGTTGAGTGGTGATTGGATTCCACTCAATTTACCGACCCAAATCCAATCATTAT 23854

Qy      2423 GGCCCGGCGTGTCGGTGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCG 2482
Db     23855 GGCCCCAAGGACAGGTGGTCAGTCTGGGGGGAGCCACAGAGGCTTCAATTTGGTCTGTGT 23914

Qy      2483 GGTACCCCGTGAGGAACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGC 2542
Db     23915 ACTATCCCATCACTACCGTAGAGCCGGAATGGAAGAGCATTCCCTACGGCAAACCTCTAG 23974

Qy      2543 GCAACCAGACGTTCCACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGG 2602
Db     23975 TGAATCAAAGCCTACATGTGCTCAATCACAATCTGGACCCTTGTCCTAATTGGGTTCCAG 24034

Qy      2603 GGCAACTCTACATTGGCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGA 2662
Db     24035 GTCAGCTCTACATTGGTGGGATTGGTCTGGCCCAGGGCTATTGGCGGGATGAGCAAAAAA 24094

Qy      2663 CGCGCAAGAGCTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATC 2722
Db     24095 CGAATGCCAGCTTTATCCTGCATCCCCAGACTGGAGAGCGCTTGTATAAAACGGGTGATT 24154

Qy      2723 TGGGCCGCTACCTGCCCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCA 2782
Db     24155 TGGCCCGCTACTTACCCGACGGCAACAGTGAATTTTTGGGTCGGGAGGATTTCCAAGTCA 24214

Qy      2783 AGCTTCGCGGATACCGCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGA 2842
Db     24215 AAATCAGTGGTTATCGAATTGAACTCGGTGAGATTGAGGCAACCCTGCTGGGGCATGCTA 24274

Qy      2843 ACGTACGCGACGCGGTGATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAG 2902
Db     24275 CAGTTAAGGAAACCG---TTGTGGCTGCTGTGGGTGAACTACAGAGTAAGCAGTTGGTCG 24331

Qy      2903 CCTATGTGGTCCCGGAGGGCACACGGAGACGCGCTGCCGAGCAGGACGCGAGCCTCAAGA 2962
Db     24332 CCTATGTGGTCTTTCACTCAGAATCATCATCTGACTCTGCCACAGAAGACGTCCACGATG 24391

Qy      2963 CCGAGCGGATCGACG 2977
Db     24392 ACATGCGCATCGACG 24406
```

```
RESULT 14
SCO939132
LOCUS       SCO939132              302007 bp    DNA     linear   BCT 11-FEB-2003
DEFINITION  Streptomyces coelicolor A3(2) complete genome; segment 29/29.
ACCESSION   AL939132 AL355753 AL391338 AL391395 AL391398 AL391454 AL446003
            AL450165 AL450223 AL450289 AL450350 AL450432 AL450433 AL512944
            AL596144 AL645882
VERSION     AL939132.1  GI:24413926
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2)
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
            Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
            Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
            Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
            Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S.,
            Rabbinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
            Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
            Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
            Parkhill,J. and Hopwood,D.A.
  TITLE     Complete genome sequence of the model actinomycete Streptomyces
            coelicolor A3(2)
  JOURNAL   Nature 417 (6885), 141-147 (2002)
  MEDLINE   21996410
   PUBMED   12000953
REFERENCE   2  (bases 1 to 302007)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
            sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT     On or before Oct 26, 2002 this sequence version replaced
            gi:20520855, gi:9843813, gi:9857138, gi:9857140, gi:20520741,
            gi:20520786, gi:20520787, gi:20520842, gi:20520681, gi:20520707,
            gi:20520788, gi:20520682, gi:20520861, gi:20520876.
FEATURES             Location/Qualifiers
     source          1..302007
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
     gene            116..1390
                     /gene="SCO7543"
                     /note="synonym: SC8G12.19"
     CDS             116..1390
                     /gene="SCO7543"
                     /note="SC8G12.19, possible Crp-family transcriptional
                     regulator, len: 424 aa. Similar to several including:
                     Bacillus subtilis SW:XYLR_BACSU (EMBL:M27248) xylose
                     repressor (384 aa), fasta scores opt: 339 z-score: 347.1
                     E(): 7.1e-12 23.3% identity in 390 aa overlap and
                     Streptomyces coelicolor TR:O50502 (EMBL:AL009199) probable
                     transcriptional repressor protein, SC7B7.05 (403 aa),
                     fasta scores opt: 691 z-score: 697.8 E(): 2.1e-31 33.9%
                     identity in 387 aa overlap. Contains a Prosite hit to
                     PS01125 ROK family signature and Pfam hits to PF00480 ROK,
                     ROK family and PF00325 crp, Bacterial regulatory proteins,
                     crp family with the putative helix-turn-helix motif
                     situated between residues 32..53 (+4.61 SD)."
                     /codon_start=1
                     /transl_table=11
                     /product="putative Crp-family transcriptional regulator"
                     /protein_id="CAB90901.1"
                     /db_xref="GI:7799279"
                     /translation="MTGMQMPSGVHALVRRTHEERVMRALQENGAMSRGEIARVVGLS
                     RTTLSEITGNLLQRGAIVVVDTDASRREGSGRPAERLALDPASAQFMGVDFGHRRVHV
                     VVADAAHEIMVSGSVRYEDTATWPTRTERALELVDRLSGEAGVHYGALQGIGIGVPGP
                     YPAPEGAAWPRATPGTTVLRPAPEGVDVAFAERFDAPVIVDNNTRLAALAEAISGADS
                     VADLVYVRLSDGVGGGLVVGGQLVTGSSGLAGELGHVTVEPAGRPCRCGKRGCLETVA
                     SVPGILAACWEFGLRLENLDDLAAAVRRAHPVVDRVLREAAGALGRVVGAATMMLNPA
                     KVVIGGEITRLAPVLVEQVAATLAAEIFPTASAGPVVAAARLSDDDGAIGALAAVFHS
                     SPLLARYPETADVKGRSDAPHSATEGAAHVRP"
     misc_feature    206..298
                     /gene="SCO7543"
                     /note="Pfam match to entry PF00325 crp, Bacterial
                     regulatory proteins, crp family, score 15.80, E-value
                     0.0014"
     misc_feature    680..958
                     /gene="SCO7543"
                     /note="Pfam match to entry PF00480 ROK, ROK family, score
                     96.80, E-value 1.1e-27"
     misc_feature    794..877
                     /gene="SCO7543"
                     /note="PS01125 ROK family signature"
     gene            1485..2288
                     /gene="SCO7544"
                     /note="synonym: SC8G12.20"
     CDS             1485..2288
                     /gene="SCO7544"
                     /note="SC8G12.20, possible ABC-transport protein, membrane
                     component, len: 267 aa. Similar to many other transporters
                     e.g. Escherichia coli SW:TAUC_ECOLI(EMBL:D85613) taurine
                     transport permease protein, TauC (275 aa), fasta scores
                     opt: 547 z-score: 643.1 E(): 2.3e-28 37.5% identity in 248
                     aa overlap. Contains multiple possible membrane spanning
                     hydrophobic domains and a Pfam match to entry PF00528
                     BPD_transp, Binding-protein-dependent transport systems
                     inner membrane component."
                     /codon_start=1
                     /transl_table=11
                     /product="putative ABC-transport protein, membrane
                     component."
                     /protein_id="CAB90902.1"
                     /db_xref="GI:7799280"
                     /translation="MAADTSAGAAESAPERSRSGRRGLPLVLNAVSVALGIALWWALA
                     SAGFKLPTPPEVVDRASTLIGDGTLADDALASLTRVLIGFGLGTAVAVPVGFLMGWYG
                     ILRGLIEPWIQFFRTIPPLAIIPLAVVAMGIDETPKIFVIFLAAFLACVISTFQGVVN
                     VDRTLINAARVLGAGDATIFARVVVPASTPFILVGMRVGLGSAWATLVAAELIAAQEG
                     LGYRMQNAQLYYDLPTIFVGLISIGILGLLMDRVLLLAERRLTRWQERR"
     misc_feature    1953..2168
                     /gene="SCO7544"
                     /note="Pfam match to entry PF00528 BPD_transp,
                     Binding-protein-dependent transport systems inner membrane
                     component, score 44.90, E-value 1.8e-09"
     gene            2285..3097
                     /gene="SCO7545"
                     /note="synonym: SC8G12.21"
     CDS             2285..3097
                     /gene="SCO7545"
                     /note="SC8G12.21, possible ABC-transport protein,
                     ATP-binding component, len: 270 aa. Similar to multiple
                     ABC-transport proteins e.g. Escherichia coli
                     SW:TAUB_ECOLI(EMBL:D85613) taurine transport ATP-binding
                     protein, TauB (255 aa), fasta scores opt: 665 z-score:
                     772.8 E(): 0 43.4% identity in 235 aa overlap. Contains a
                     Prosite hit to PS00017 ATP/GTP-binding site motif A
                     (P-loop) and a Pfam match to entry PF00005 ABC_tran, ABC
                     transporter."
                     /codon_start=1
                     /transl_table=11
                     /product="putative ABC-transport protein, ATP-binding
                     component."
                     /protein_id="CAB90903.1"
                     /db_xref="GI:7799281"
                     /translation="MTSTPAKISVQHVTKTFSLGRETFTALDDVSLDIANHEFVTVVG
                     PSGCGKSTLMNILAGLEAPTSGRALVDGAPVSGPGPERGVIFQQYALFPWLTVRQNVE
                     FGLRTTGVPKRERRARAEHFIELVGLERFADALPKTLSGGMRQRCAIARAYAVDPSIL
                     LMDEPFGALDALTRVKLQEQLLKTWSQDRRTVLFITHDVDEAVFLANRVVVMAARPGR
                     VYDVVDVDPAIERDEAFRLSPEFAALRNRVWRSVYHQEGRTAALSGSHSPTP"
     misc_feature    2393..2929
                     /gene="SCO7545"
                     /note="Pfam match to entry PF00005 ABC_tran, ABC
                     transporter, score 224.10, E-value 2e-63"
```

```
     misc_feature    2414. .2437
                     /gene="SCO7545"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     gene            3094. .4116
                     /gene="SCO7546"
                     /note="synonym: SC8G12.22"
     CDS             3094. .4116
                     /gene="SCO7546"
                     /note="SC8G12.22, possible solute-binding lipoprotein,
                     len: 340 aa. Similar to several e.g. Escherichia coli
                     SW:TAUA_ECOLI(EMBL:D85613) taurine-binding periplasmic
                     protein precursor, TauA (320 aa), fasta scores opt: 235
                     z-score: 267.4 E(): 1.9e-07 23.3% identity in 287 aa
                     overlap. Contains a possible N-terminal signal sequence
                     and an appropriately positioned Prosite hit to PS00013
                     Prokaryotic membrane lipoprotein lipid attachment site."
                     /codon_start=1
                     /transl_table=11
                     /product="putative lipoprotein."
                     /protein_id="CAB90904.1"
                     /db_xref="GI:7799282"
                     /translation="MTRKGPTAMSPVLTRRSLTTTLAGVAAMVLLAACGGSSSGGSSD
                     KVRFGYINDFNGASLIAIAEAEGLWKKHGLSADAKVFTNGPLQIQALGTDNLDFGYIG
                     PGAMWLPASGQAKVVAINTLGNSDRVLAQPGITSMQQLKGKTVAVPEGTSGDMILSLA
                     LEKAGMTKEDLKVVPMDPSTIVSAFSSQQVDGAGFWYPAAATIKKQVPDLVELAKNAD
                     FEQDVSFPTAFVAGNKVVADQPEKTKKVLAVLREAMTFRSEHTDEAIKLTADKLKIPV
                     EQVKADAANNKVLSVAELDELTRDGTVVKWLEGMNDYFVNAGKLKKPADPKTYYTGDL
                     FTGAGK"
     misc_feature    3163. .3195
                     /gene="SCO7546"
                     /note="PS00013 Prokaryotic membrane lipoprotein lipid
                     attachment site"
     gene            4113. .5714
                     /gene="SCO7547"
                     /note="synonyms: SC5F1.01, SC8G12.23"
     CDS             4113. .5714
                     /gene="SCO7547"
                     /note="SC5F1.01, possible sulfatase (fragment), len: >136
                     aa; low similarity to SW:YIDJ_ECOLI (EMBL:L10328)
                     Escherichia coli hypothetical 57.3 kDa protein in
                     EmrD-GlvG intergenic region putative sulfatase YdjI (EC
                     3.1.6.-), 497 aa; fasta scores: opt: 147 z-score: 185.4
                     E(): 0.0085; 29.6% identity in 115 aa overlap
                     SC8G12.23, possible sulphatase, partial CDS, len: >430 aa.
                     Similar to many Eukaryotic sulphatases e.g. Homo sapiens
                     (Human) SW:IDS_HUMAN(EMBL:M58342) iduronate 2-sulfatase
                     precursor (EC 3.1.6.13) (550 aa), fasta scores opt: 326
                     z-score: 382.2 E(): 7.9e-14 28.4% identity in 423 aa
                     overlap. Also similar to Escherichia coli
                     TR:YIDJ_ECOLI(EMBL:L10328) hypothetical 57.3 kd protein
                     (497 aa), fasta scores opt: 221 z-score:0.0 E(): 0.0
                     31.23% identity in 397 aa overlap. Contains a Prosite hit
                     to PS00523 Sulfatases signature 1 and a Pfam match to
                     entry PF00884 Sulfatase, Sulfatase."
                     /codon_start=1
                     /transl_table=11

  Query Match             14.2%;  Score 600;  DB 1;  Length 302007;
  Best Local Similarity   52.8%;  Pred. No. 3.9e-85;
  Matches 1622;  Conservative    0;  Mismatches 1350;  Indels  102;  Gaps   11;

Qy          2 TGACGATCAATCAGCTTCTGAACGAGCTCGAGCACCAGGGTATCAAGCTGGCGGCCGATG 61

Db     147470 TGAACATCGCTGAACTGCTGGCCCGATACGCGGACGAAGGCGTGGTGCTGTGGGCCGAGG 147529

Qy         62 GGGAGCGCCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCCGAACCTGCTCGCTCGAA 121

Db     147530 ACGGACAACTGCGGTTCCGGGCCCCGCAGGGCCGGCTCACCGACACGCACCGAGCCGAAC 147589

Qy        122 TCTCCGAGCACAAAAGCACGATCCTGACGATGCTCCGTCAGAGACTCCCCGCAGAATCCA 181

Db     147590 TGCGCACGCACAAGGAAGCCGTCC------TGCGCCATCTGGAAGCGGCGGAACCGGCGC 147643

Qy        182 TCGTGCCCGCCCCAGCCGAGCGGCACGCTCCGTTTCCTCTCACAGACATCCAAGAATCCT 241

Db     147644 TGCGGGCCGACCCCGACCACCGCCACGAACCCTTCCCCCTGACGGACATCCAGGCCGCCT 147703

Qy        242 ACTGGCTGGGCCGGACAGGAGCGTTTAC---GGTCCCCAGCGGGATCCACGCCTATCGCG 298

Db     147704 ACCTCATCGGCCGCACCGACGCCTACGCCTACGGGGGCGTCGGCTGCCACGCCTACGTGG 147763

Qy        299 AATACGACTGTACGGATCTCGACGTGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCG 358

Db     147764 AGCTGGCATATCCCGACCTCGACGTCGACCGAGTCACCTCCGTGTGGCGGCAACTGGTCC 147823

Qy        359 CGCGGCACGACATGCTTCGGGCCCACACGCTGCCCGACATGATGCAGGTGATCGAGCCTA 418

Db     147824 AGCGGCACGACATGCTCCGGGCCGTCATCCACCAGGACGGCCACCAGCGGGTCCTGCCCG 147883

Qy        419 AAGTCGACGC---CGACATCGAGATCATCGATCTGCGCGGGCTCGACCGGAGCACACGGG 475

Db     147884 AGGTGCCCGCGCTGTCCGTCGGGACCGACGACCTGACCGCCCTGCCGGCCGCCACCGCCG 147943

Qy        476 AAGCGAGGCTCGTGTCGTTGCGAGATGCGATGTCGCACCGCATCTATGACACCGAGCGCC 535

Db     147944 AGGCACGGATGCGGCACACCCGGGCGCGGCTGTCCGCCCGCGAGGCGCCCACCGACCAGT 148003

Qy        536 CTCCGCTCTATCACGTCGTCGCCGTTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCA 595

Db     148004 GGCCCCTCT---TCGACGTACACGTGACCCGGACCGACCGGCGGGCCGTACTGCACCTCT 148060

Qy        596 GTATCGATCTCATTAACGTTGACCTAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCA 655

Db     148061 CCTTCGACATGCTCGTCGTCGACCACGCCAGCCTGCGGATCCTGCTCGCCGAGTTCCGGC 148120

Qy        656 GCTTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATG 715

Db     148121 GGAGCTACGGCGGCACCG---CCCTCACCGACGCGCCGCGGATCACTTTCCGCGACTACG 148177

Qy        716 TACTCGCGCTGGAGTCTCGCAAGAAGTCTGAGGCGCATCAACGATCGATGGATTACTGGA 775

Db     148178 TCCTGGCCCGCCGCGCCCTTGCCGACACCGACGGCCACGCCCGCGACCGCGCCTACTGGA 148237

Qy        776 AGCGGCGCATCGCCGAGCTCCCACCTCC-------------------------------- 803

Db     148238 CCGAACGGCTCGACACCCTGCCGCCCGCCCCCGAACTGCCCCTCGCCGAAGCCTGGCAGG 148297

Qy        804 --------GCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTGAAGGAGATCCG---- 851

Db     148298 CCGCGGTGGACGACCCGGCGGAAGCGGGCCGCGTCCCGGCGCCCGGCGGCGACGCGGTGG 148357

Qy        852 -CTTCCGGCACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTGAAGCGGCGTG 910

Db     148358 CGTTCCGCCGGCTGGAGGTGCTGCTGCCCGCGGCCGACCGGGACCGGCTCACCGCACGTG 148417

Qy        911 TCGGGGAGCGCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCCGAGGTGATCG 970

Db     148418 CCGCGCGGCGCGGGCTGACCCCGTCGACGGCCCTGCTCACCGCTTACGCCGAGACGGTCG 148477

Qy        971 GGCGCTGGAGCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAACCGGCTCCCCG 1030

Db     148478 GCGCCTGGTCGCGCTCCAGCCGGTTCACCCTCAACGTGCCCACCGTCGACCGGCCCGCCC 148537

Qy       1031 TCCATCCGCGCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTCCTGGACATCG 1090

Db     148538 TGCACGAGGACATCGACCGCCTCGTCGGCGACTTCACCTCGCTCGAACTCCTCGCCGTCG 148597

Qy       1091 ACACCACTCGCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAGCAGCTGTGGG 1150

Db     148598 ACCTCGACACCCCGGCCACCTTCGCCGAACGGACCGGGGCCGTGGGCGAACAGCTCCTCG 148657

Qy       1151 AAGCGATGGATCACTGGCACGTAAGCGGTATCGAGGTCCAGCGAGAGGCCGCCCGGGTCC 1210

Db     148658 ATGACCTGGCCCACCCGCTGTTCACCGGCTCGGAGGTGC------TCGCCGAACTCTCCC 148711

Qy       1211 TGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTTAACCAGCAAG 1270
```

```
Db     148712 GCAGGGCCGGCGCTCCGGTCCTCATGCCGGTCGTCTTCACCAGCGCCCTG---------G 148762

Qy       1271 TCGTTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGCACGCAGACTC 1330
Db     148763 GCGCGGGTGCCACCTCCGAGGGCGTACCCCCGAGGTGGAGTACGCCGTGACCCGGACAC 148822

Qy       1331 CTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTCCTCGCGTGGG 1390
Db     148823 CGCAGGTGTGGCTCGACTGCCAGGTCATGCACCGCGGCGACACCCTCAGCCTGTCCTGGG 148882

Qy       1391 ACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAAGCGTACGTCG 1450
Db     148883 ACATCCGGGAGGGCGCGCTCGCCCACGGGACGGCCGACGCCATGTTCGAGGCGTACACGG 148942

Qy       1451 TTTTTCTCCGGCGGCTCACT-----------------GAGGAACCATGGGGTGAACAGG 1492
Db     148943 CACTCGTGCGTTCGCTGTCGGCCGAGGGCGAGGCAGGCGAGAAGGCCTGGGACGCACCCG 149002

Qy       1493 TGCGCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACG 1552
Db     149003 TCCGCATCCCGCTGCCCGCGGCACAGGCGGCACGCCGGGCCGCGGTCAACGCGACCGAGG 149062

Qy       1553 CGCTGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCA 1612
Db     149063 GCCCCCTCCCGGACGCCCTGCTCCACGAACCCGTACTCGCCCGGGCACGGACCACACCGG 149122

Qy       1613 TGCAGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTT 1672
Db     149123 ACGCGATCGCGGTCCGCACCCCCGAACTGGCCCTCTCCTACCGGCAGCTCGTCACCCGCG 149182

Qy       1673 CGCGGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGG 1732
Db     149183 CCACAGGACTCGCCCAGCACCTGACGGCCTCCGGACTGCGCCCCGGGGAACCGGTGGCCA 149242

Qy       1733 TGGTGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCG 1792
Db     149243 TCTGGATGGACAAGGGCTGGGAACAGGTCGTCGCCGTCTTCGGCACCCTCATGGCGGGCG 149302

Qy       1793 CGGCCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATC 1852
Db     149303 GCGCCTACCTGCCCGTCGACACGGCCCAGCCGGCGGCCCGCCGGGACACGATCATCGGGG 149362

Qy       1853 ATGGTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGC 1912
Db     149363 ACGCGGGCGTCCGCACCGTCCTCACCCAGTCCTGGCTCGCCGAACTCGAAGACCTCCCCT 149422

Qy       1913 CGGGGATCCAGCGGCTGCTCGTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGCCTC 1972
Db     149423 CGACGGTGTCGCCGGTCGCCGT------GGACCTCGTCGGGGAGGCCACCGCGGACCTCC 149476

Qy       1973 CGATGATGCCCATTCAGACACCTTCGGATCTCGCGTATGTCATCTACACCTCGGGATCCA 2032
Db     149477 CGCCCGCCGCCCGCCGGGACCCTGACGACCTCGCCTACGTCATCTACACCTCGGGCTCCA 149536

Qy       2033 CAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCCTGGACA 2092
Db     149537 CCGGCACGCCCAAAGGCGTGATGATCAGCCACCGGGCCGCCCTCAACACCGTCGAGGACA 149596

Qy       2093 TCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGCTCTCCTCGCTGAGCT 2152
Db     149597 TCAACCGCCGCTTCGCCGTCGACGAGCGCGACCGCGTCCTCGGCATCGCCGGACTCGGCT 149656

Qy       2153 TCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCGGTACGATCGTGGTGC 2212
Db     149657 TCGACCTGTCCGTCTACGACCTGTTCGGACCGCTCGCCGTGGGCGCCACCCTCGTCCTGC 149716

Qy       2213 CGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGATCGAACGAGAGAAGG 2272
Db     149717 CCCGGTCCGACCTGCGCGGCGACCCCTCCCACTGGGCCGAACTCGTCCGCGACTTCGGCG 149776

Qy       2273 TGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCGAGCATTTTGAGGGTC 2332
Db     149777 TCACGGTGTGGAACTCCGTACCGGGCCAGCTCCACATGCTGTGCGACTGGCTCCGCTCCG 149836

Qy       2333 GCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGG 2392
Db     149837 AGCCCCCCACCGACGACGGCTCCCTGCGCCTGGCACTGATATCGGGCGACTGGATACCGG 149896

Qy       2393 TGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGGTGATCAGCCTGGGCG 2452
Db     149897 TCGCACTGCCCGACCAGGCCCGTGAACTCCTGCCCGGACTGGAGATCGTCTCCCTCGGCG 149956

Qy       2453 GGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGAACGTCGACCTATCGT 2512
Db     149957 GAGCCACCGAGGGCTCCATCTGGTCCATCGCCCACCCGATCGGCGAGGTCGACACCGCCC 150016

Qy       2513 GGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCCACGTGCTCGATGAGG 2572
Db     150017 GCCCCAGCATCCCGTACGGCAAGCCCCTGACCAACCAGACGTTCGCCGTGCTCGACCGGC 150076

Qy       2573 CGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTGGCGGGGTCGGGCTGG 2632
Db     150077 ACCTGCGGCCGCGGCCCGAGTGGGTGCCGGGCGAGCTGTACATCGGCGGTGCCGGTGTCG 150136

Qy       2633 CACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCCTCGTGCACCCCGAGA 2692
Db     150137 CCCTGGGCTACCTCGGCGACGGCGAGCGCACGGCACAACGGTTCCTCACCGACTTGGCCA 150196

Qy       2693 CCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGCCCGATGGAAACATCG 2752
Db     150197 CCGGAGAGCGGCTCTACCGCACGGGCGACCTCGGCCGCTACCTGCCCGACGGGACCATCG 150256

Qy       2753 AGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACCGCGTTGAGCTCGGGG 2812
Db     150257 AGTTCCTGGGCCGGGAGGACGCCCAGATCAAGATCCGCGGATACCGGGTCGAACTCGCCG 150316

Qy       2813 AAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGGTGATTGTGCCCGTCG 2872
Db     150317 AGGTGGAGGCCGCCGTACAGACCCACCCCGCGGTCGCCGCGGGCGCGGTCGTCGTGGACG 150376

Qy       2873 GGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGGAGGGCACACGGAGAC 2932
Db     150377 ACTCGGCGGCGGGCGGGCGGCGCCTGGCCGCGTTCGTCGAGACGGCACGCAAGGACGGCG 150436

Qy       2933 GCGCTGCCGAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACGCGAGAGCACACGCCG 2992
Db     150437 GGCCCGCCGTGACGAGCCGCGCACAGGCGCGCACGGCAGCCGCCGAAGCGGTACGCGAAG 150496

Qy       2993 CCGAAGCGGACGGC 3006
Db     150497 CGGGCGCGGCCGTC 150510
```

RESULT 15
AL646066

```
LOCUS       AL646066               189050 bp    DNA     linear   BCT 11-APR-2003
DEFINITION  Ralstonia solanacearum GMI1000 chromosome, complete sequence;
            segment 10/19.
ACCESSION   AL646066 AL646052
VERSION     AL646066.1  GI:17428714
KEYWORDS    .
SOURCE      Ralstonia solanacearum
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
            Burkholderiaceae; Ralstonia.
REFERENCE   1
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Nature 415 (6871), 497-502 (2002)
  MEDLINE   21681879
```

```
  PUBMED    11823852
REFERENCE   2  (bases 1 to 189050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .189050
                     /organism="Ralstonia solanacearum"
                     /mol_type="genomic DNA"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
     gene            170. .3325
                     /gene="RSc1698"
                     /note="synonym: RS02885"
     CDS             170. .3325
                     /gene="RSc1698"
                     /function="elements of external origin; phage-related
                     functions and prophages"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE BACTERIOPHAGE PROTEIN"
                     /protein_id="CAD15400.1"
                     /db_xref="GI:17428715"
                     /db_xref="SPTREMBL:Q8XYR0"
                     /translation="MRNIIGYGGGKGGSSSSTPTEAPDSLHSVAYARVLDLVSEGEIA
                     GLVNGLQSIYLEGTPLANADGTLNFQNVAVDYRPGSQDQEAIPGFPSVESESAVGVEL
                     TATMPWVRAIANTQLSAVRVQLSVPALSRADTSSGNINGYRVEYAIDLAVDGGALQQV
                     LATAFDGKTTSKYTRTHRIELPRAKTGWTIRVRRITPNANSGTIADVTRIESIAEVID
                     AKLRYPNSALIGVRIDARQFSSVPTRSYHLRGRIIRVPSNYDPATRTYTGLWDGTFKV
                     AYSNNPAWVFYDIVLHKRCGLGDRVSADMVDKWALYQIGQYCDELVPDGRGGQEPRFT
                     CNCYLQQRNDAYPVLQDLASVFRGMAYWASSNVVAVADMPASASYLFHPGNVVDGRFT
                     YAGSARRARKTVALVSWNDPGDQYRAKVEHVEDGDGIARYSIQQTEVTAFGCTSQSQA
                     QRVGQWILLTSRLETETVSFKVGLDAAVVMPGAIIEVADPARAGRTNGGRVRSADGRT
                     VVVDRPVVVREGDTLLLNLPDGTAQRRTISRVDGQSLTVSADWSQPVQAEAVWSVESA
                     DLKTQLFRVVSVTESDGLTFDISALEYNQSKFSAIDNGTRIEARPVSVLPPSVQPAPT
                     DVTLATYSVVDQGIAITTMMVAWQQAASAVAYTVEWRRDNGEWVTAGRTGSQSLEVRN
                     IYAGTYVARVRAINALDVASLPAYSAETRLTGKTSPPPVVGTLIATAIVFGIRLDWAF
                     PTGPLDVERTELWWSKAPDRASATKLGDFAFPANTHTLMGLAAGATFYFWARLVDKSG
                     NVGAWYPSGNGVLGMSSDRASDILSYMKGQIGKTQLAGDLLSAVDSIDPPMAGSDEDF
                     AGDGTVYAGIVSTQSVLEEANRAIAQQVTTMQATVAQSSAAVQVAQQAVAEQGGRLAA
                     MYTIKTQIAANGRIYLAGIGFGVENDKGVIESQVLIAADRFAVLHPNGNNVFTPFVVQ
                     GGQVFMDSAFIADGTITSAKIGDSIQSSNFVAGQTGWRLSKSGLFENNGNGAGGRRVD
                     TSALTQIYDANGTLRVRLGVW"
     gene            3662. .4318
                     /gene="RSc1699"
                     /note="synonym: RS02888"
     CDS             3662. .4318
                     /gene="RSc1699"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAD15401.1"
                     /db_xref="GI:17428716"
                     /db_xref="SPTREMBL:Q8XYQ9"
                     /translation="MPAGLTVYGEHGFVQITEAYGNLALRQKSTVTPDGSGRGSITFS
                     ASRPFVCIQSANPVALLGSSNSGADWTVKFSAPAQAPFTVYVFDEPAAPSGQPGLQVF
                     KADGTLAFDSGLAYLKVSGVVTPPSGAPAYGQAWEASPLTAGNYAACLSYTRTGIYAV
                     PGTDTLFVADHVYTTANGAGLKLLQYTHRGAWNQGDGSGLKIDTANPPQIVLVDVSQL
                     "
     gene            4387. .4878
                     /gene="RSc1700"
                     /note="synonym: RS02889"
     CDS             4387. .4878
                     /gene="RSc1700"
                     /EC_number="3.2.1.-"
                     /function="elements of external origin; phage-related
                     functions and prophages"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE BACTERIOPHAGE ENDOLYSIN PROTEIN"
                     /protein_id="CAD15402.1"
                     /db_xref="GI:17428717"
                     /db_xref="GOA:Q8XYQ8"
                     /db_xref="SPTREMBL:Q8XYQ8"
                     /translation="MPYTDPSVLGGRNLAAFLDMVGFSEGTDNGRQPTRDRGYDVLVG
                     GGLFVSYADHPRILVELPRLGIKSTAAGRYQLLARYYDPYRRQLRLVDFGPAAQDAIA
                     VQQIRERGALADIQAGRLAAAIAKCKNIWASLPGAGYGQHEQKFEALRAHYLLCGGQE
                     VKG"
     gene            4878. .5186
                     /gene="RSc1701"
                     /note="synonym: RS02890"
     CDS             4878. .5186
                     /gene="RSc1701"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE LIPOPROTEIN TRANSMEMBRANE"
                     /protein_id="CAD15403.1"
                     /db_xref="GI:17428718"
                     /db_xref="SPTREMBL:Q8XYQ7"
                     /translation="MTESEMAVAAKVGGSAALGSVIALRFLPGTWWQRMLSFVSSLGI
                     GCLAGGAAVERFLLVPGSYTHMLAVASAAVFGLAIVNNAMQQIPEILTDLRRRFLAKE
                     "
     gene            5189. .5464
                     /gene="RSc1702"
                     /note="synonym: RS02891"
     CDS             5189. .5464
                     /gene="RSc1702"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE TRANSMEMBRANE PROTEIN"
                     /protein_id="CAD15404.1"
                     /db_xref="GI:17428719"
                     /db_xref="SPTREMBL:Q8XYQ6"
                     /translation="MLLTSINQVANAIIFVGSLWAVLTHQVPTRTGGALVLALVNFAA
                     LGNIVAHGTCHSWPEVALNVAVALCVVWAFWRLEVRRRFSALGKGRG"
```

```
gene            5464..6009
                /gene="RSc1703"
                /note="synonym: RS02892"
CDS             5464..6009
                /gene="RSc1703"
                /EC_number="3.4.-.-"
                /function="elements of external origin; phage-related
                functions and prophages"
                /note="Product confidence : probable
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PROBABLE BACTERIOPHAGE-RELATED SIGNAL PEPTIDE
                PROTEIN"
                /protein_id="CAD15405.1"
                /db_xref="GI:17428720"
                /db_xref="GOA:Q8XYQ5"
                /db_xref="SPTREMBL:Q8XYQ5"
                /translation="MPGIKGTAIAAAIAMLLGALVGAAGAWTWQAHRYDKDIATLQRD
                QAEKDRQQARADEAQLRQEREKRAGLERQLGELDARHYEELNNAKTENARLAAALAAG
                TRQLSIRARCTAPADGVPSTTSGAGLGDGGAATAILYGEDAADIQALVADADAIAAAL
                RFFQGRERAIQTRDTSVDARR"
repeat_region   complement(6117..7304)
                /note="predicted by Homology"
                /insertion_seq="ISRso18 or RS05932"
gene            complement(6273..7238)
                /gene="tISRso18"
                /note="synonyms: RS02893, RSc1704"
CDS             complement(6273..7238)
                /gene="tISRso18"
                /function="elements of external origin; transposon-related

Query Match             14.0%;  Score 593.6;  DB 1;  Length 189050;
Best Local Similarity   52.6%;  Pred. No. 4.3e-84;
Matches 1510;  Conservative    0;  Mismatches 1309;  Indels   51;  Gaps    8;

Qy        206 ACGCTCCGTTTCCTCTCACAGACATCCAAGAATCCTACTGGCTGGGCCGGACAGGAGCGT 265
Db     159271 ACGACACCTTCGCCCTGACCGACATCCAGCACGCTTACTGGATCGGCCGCGGCCAGGCCC 159330

Qy        266 TTACGGTCCCCAGCGGG---ATCCACGCCTATCGCGAATACGACTGTACGGATCTCGACG 322
Db     159331 TGGCGCTCGGCGGCGTGTCATGCCACGTCTACTTCGAGTGGGTCTCGCCAGACCTGGACC 159390

Qy        323 TGCCGAGGCTGAGCCGCGCCTTTCGGAAAGTCGTCGCGCGGCACGACATGCTTCGGGCCC 382
Db     159391 TGCAACGCTTCGAGCAGGCCTGGAACCGGATCGTGGCGCGGCACGGCATGATGCGCGCCG 159450

Qy        383 ACACGCTGCCCGACATGATGCAGGTGATCGAGCCTAAAGT---CGACGCCGACATCGAGA 439
Db     159451 TGCTGACGCCGGATGGGCGGCAGCGCATCCTCGGCACCGTGCCGCACTACGCCATCGAGA 159510

Qy        440 TCATCGATCTGCGCGGGCTCGACCGGAGCACACGGGAAGCGAGGCTCGTGTCGTTGCGAG 499
Db     159511 CCGAAGACCTGCGCGCGCAGCCGCCCGATCGCGTGGCGGCCAGGCTCGCGCAGAAGCGCG 159570

Qy        500 ATGCGATGTCGCACCGCATCTATGACACCGAGCGCCCTCCGCTCTATCACGTCGTCGCCG 559
Db     159571 CCGGCATGTGCGGGCAGGTGCTCGACGCCGAGCGCTGGCCGCTGTTCGACCTGCGCGCCA 159630

Qy        560 TTCGGCTGGACGAGCGGCAAACCCGTCTCGTGCTCAGTATCGATCTCATTAACGTTGACC 619
Db     159631 CGCGCGAGCCGGCCGGCGTGACCCGCCTGCACCTCGATCTCGACCTGCTGATGTTCGACG 159690

Qy        620 TAGGCAGCCTGTCCATCATCTTCAAGGACTGGCTCAGCTTCTACGAAGATCCCGAGACCT 679
Db     159691 TGCAGAGCTTCCACATCCTGCTGGCCGAGCTCGAGCGCAGCTACCGCGACCCGGCCGCGC 159750

Qy        680 CTCTCCCTGTCCTGGAGCTCTCGTACCGCGATTATGTACTCGCGCTGGAGTCTCGCAAGA 739
Db     159751 AGCTGCCCGCCATCGGCCTCAGTTTCCGGGACTACCTGCTGGCCGGCCGGGACGCCGCGG 159810

Qy        740 AGTCTGAGGCGCATCAACGATCGATGGATTACTGGAAGCGGCGCATCGCCGAGCTCCCAC 799
Db     159811 GCGAAGCAGCCTATCTGGCCGACAAGGCCTGGTGGCAGGCCCGCCTGCCAGAGATCGCGC 159870

Qy        800 CTCCGCCGACGCTTCCGATGAAGGCCGATCCATCTACCCTGAAGGAGATCCGCTTCCGGC 859
Db     159871 CGCCCCCGCGGCTGCCGCTCGCGCGCGTGGCGTCCGACGTGGCGCGGCCGACATTCCGGC 159930

Qy        860 ACACGGAGCAATGGCTGCCGTCGGACTCCTGGGGTCGATTGAAGCGGCGTGTCGGGGAGC 919
Db     159931 GCCTGCAGCATCGCTTCGACCCACAGGCGTGGCAGCGGCTGTCGGCCTTTGCCGCCGAGG 159990

Qy        920 GCGGGCTGACCCCGACGGGCGTCATCCTGGCTGCATTTTCCGAGGTGATCGGGCGCTGGA 979
Db     159991 CCGGGCTGACCGGCTCGTCCGTGCTGCTCACCGCGTTCTCGGAAGTGCTCGCGCGCTGGG 160050

Qy        980 GCGCGAGCCCCCGGTTTACGCTCAACATAACGCTCTTCAACCGGCTCCCCGTCCATCCGC 1039
Db     160051 CCGGCGACGGCCGCTTCACGCTCAACCTCACGCACTTCAACCGCCGCCCGCGTGCATCCGG 160110

Qy       1040 GCGTGAACGATATCACCGGGGACTTCACGTCGATGGTCCTCCTGGACATCGACACCACTC 1099
Db     160111 ATGTCGAGCGGCTGGTCGGCGATTTCACCTCGGTGCTGCTGCTCACGGTCGACTGCAACG 160170

Qy       1100 GCGACAAGAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAGCAGCTGTGGGAAGCGATGG 1159
Db     160171 CGCCGCTGGCCTTCGCCCAGCGTGCCGCCGCCACCCAGGCCCAGCTCTGGGCCAGCCTGT 160230

Qy       1160 ATCACTGCGACGTAAGCGGTATCGAGGTCCAGCGAGAGGCCGCCCG-------------- 1205
Db     160231 CGCACACGCGCTTCGGCGGCGTGGAAGTCCTGCGCGAACTGGCCCGCCACCGCGCGCAGG 160290

Qy       1206 -GGTCCTGGGGATCCAACGAGGCGCATTGTTCCCCGTGGTGCTCACGAGCGCGCTTAACC 1264
Db     160291 ACGGCGGCGGCATGGACGGCGGCAGCCTGATGCCGGTCGTGTTCACCAGCCTCCTGGGCA 160350

Qy       1265 AGCAAGTCG---TTGGTGTCACCTCGTTGCAGAGGCTCGGAACTCCGGTGTACACCAGCA 1321
Db     160351 TGGATCTCGACCTGCTGGTCCGTGGCGCGGACCTGCTCGGCGAGCCGGAATTCCTCTACA 160410

Qy       1322 CGCAGACTCCTCAGCTGCTGCTGGATCATCAGCTCTACGAGCACGATGGGGACCTCGTCC 1381
Db     160411 CCTGCACGCCCCAGGTCTGGCTCGACCACCAGGTGATGGTCCGCAAGGGCACGCTGGAAT 160470

Qy       1382 TCGCGTGGGACATCGTCGACGGAGTGTTCCCGCCCGACCTTCTGGACGACATGCTCGAAG 1441
Db     160471 TCAACTGGATCGTGATCGACGACCTGTTCCCGGAGGGCATGGTCGATGCGATGTTCGAGG 160530

Qy       1442 CGTACGTCGTTTTTCTCCGGCGGCTCACTGAGGA------ACCATGGGGTGAACAGGTGC 1495
Db     160531 CCTATTGCCGGCGGCTCGACCATCTGGCCCAGGATGCCGCGCATTGGCAGCAGCCGATGC 160590

Qy       1496 GCTGTTCGCTTCCGCCTGCCCAGCTAGAAGCGCGGGCGAGCGCAAACGCGACCAACGCGC 1555
Db     160591 CTGCTGAACTGCCCGCAGCGCAGCGCGCGGTGCGCGAGGCCGTCAACCGGACCGCGCTAC 160650

Qy       1556 TGCTGAGCGAGCATACGCTGCACGGCCTGTTCGCGGCGCGGGTCGAGCAGCTGCCCATGC 1615
Db     160651 CGCTGCCGCAGTCGCCGCTGCACGCGGGCTTCTTCCGGCAGGCGGTGGCGCAGCCGCACG 160710

Qy       1616 AGCTCGCCGTGGTGTCGGCGCGCAAGACGCTCACGTACGAAGAGCTTTCGCGCCGTTCGC 1675
Db     160711 CGCTGGCCTGCGCCACCGCCCATGCCAGCGAGACCTATGCCGGACTGGCCGCCCGCGCGC 160770

Qy       1676 GGCGACTTGGCGCGCGGCTGCGCGAGCAGGGGGCACGCCCGAACACATTGGTCGCGGTGG 1735
Db     160771 TGGCGCACGTGCGCACGTTGCGCGAGGCGGGCGTCATGCCCGGCGATCTCGTTGCCTGCG 160830

Qy       1736 TGATGGAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGG 1795
Db     160831 CCATGCCCAAGGGCATTCCCGCGCTCGCCGCCGCCATCGGCATCGTCGCGGCGGGCGCGG 160890
```

```
Qy       1796 CCTACGTGCCGATCGATGCCGACCTACCGGCGGAGCGTATCCACTACCTCCTCGATCATG 1855
Db     160891 TGCTGGTGCCGTTGGCCAGCGACATGCCCCCGGCGCGCATGCAAGCCATTCTCGATGGCG 160950

Qy       1856 GTGAGGTAAAGCTCGTGCTGACGCAGCCATGGCTGGATGGCAAACTGTCATGGCCGCCGG 1915
Db     160951 CAGGTGTCAAAGCCGCCTTTATCGCGCAGGCGGACTTCACGCGGCCGGCATGCGCCAGCC 161010

Qy       1916 GGATCCAGCGGCTGCTC------GTGAGCGAGGCCGGCGTCGAAGGCGACGGCGACCAGC 1969
Db     161011 CCGTCGCGTGGCTGGTGTCGCCGGAGTCCGAGCCGGCCGTGTACGCAAGCGCCGCGCTGC 161070

Qy       1970 CTCCGATGATGCCCATTCAGACACCTTCG-----------GATCTCGCGTATGTCATCT 2017
Db     161071 AGCGGGAGATCGACGCGCACGTCGCGTCGCCCGCCCTCGATGCGCTGGCCTACATCATCT 161130

Qy       2018 ACACCTCGGGATCCACAGGGTTGCCCAAGGGGGTGATGATCGATCATCGGGGTGCCGTCA 2077
Db     161131 ATACCTCGGGCTCGACCGGCACGCCCAAGGGCGTGGCGATCACGCACGGCGCGGCCGGCA 161190

Qy       2078 ACACCATCCTGGACATCAACGAGCGCTTCGAAATAGGGCCCGGAGACAGGGTGCTGGCGC 2137
Db     161191 ACACGCTGGAAGCTGTCCACGCGCTGCTGGGCCTGACGCCGGACGACCGCGTGTTCGGCA 161250

Qy       2138 TCTCCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCCTGGCGGCGGGCG 2197
Db     161251 TCTCGGCGCTGGGCTTCGACCTGGCCATCTACGATGTCTTCGGCATGCTGTCGCGCGGCG 161310

Qy       2198 GTACGATCGTGGTGCCGGACGCGTCCAAGCTGCGCGATCCGGCGCATTGGGCAGAGTTGA 2257
Db     161311 GCGCCGTGATCCTCCCGGACGAGGCGGGCATGCGCGATGCTGCGCACTGGCTGGACCTGG 161370

Qy       2258 TCGAACGAGAGAAGGTGACGGTGTGGAACTCGGTGCCGGCGCTGATGCGGATGCTCGTCG 2317
Db     161371 CGCACCGGTACGGCGTCACGGTCTGGAACTCGGTGCCGGCCCTGCTGGCGATGCTGGTCG 161430

Qy       2318 AGCATTTTGAGGGTCGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTCGCTGCTGAGCG 2377
Db     161431 AGCGCGCGGCATCCAGCGGTCGGGGCCTGCCGGCATCGTTGCGCTGGGTCATGCTCAGCG 161490

Qy       2378 GCGACTGGATCCCGGTGGGCCTGCCTGGCGAGCTCCAGGCCATCAGGCCCGGCGTGTCGG 2437
Db     161491 GCGACTGGATTGCGCTGGACCTGCCGGCACGCCTGCGCGCACTGGCGCCGCACACGCGCC 161550

Qy       2438 TGATCAGCCTGGGCGGGGCCACCGAAGCGTCGATCTGGTCCATCGGGTACCCCGTGAGGA 2497
Db     161551 TGGCCGCCCTCGGCGGCGCCACCGAGGCCGCGATCTGGTCGAACTGGTTCGAGGTCGGGC 161610

Qy       2498 ACGTCGACCTATCGTGGGCGAGCATCCCCTACGGCCGTCCGCTGCGCAACCAGACGTTCC 2557
Db     161611 GGGTCGATCCCGCCTGGCGCTCCATACCGTACGGGTTTCCGCTGGCCAACCAGTCCTACC 161670

Qy       2558 ACGTGCTCGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGCAACTCTACATTG 2617
Db     161671 GGGTGCTCGATGCCCACCTGCGCGATTGCCCGGACTGGGTCGATGGCGATTTGCATATCG 161730

Qy       2618 GCGGGGTCGGGCTGGCACTGGGCTACTGGCGCGATGAAGAGAAGACGCGCAAGAGCTTCC 2677
Db     161731 GCGGCGCCGGCCTGGCTTCGTGCTACTGGCGCGATCCGCAGCGCACCGCCGAGGCCTTCA 161790

Qy       2678 TCGTGCACCCCGAGACCGGGGAGCGCCTCTACAAGACCGGCGATCTGGGCCGCTACCTGC 2737
Db     161791 TCACCCACCCCGCCAGCGCAGAGCGCCTGTACCGCACCGGCGACCGCGCGCGCTACTGGC 161850

Qy       2738 CCGATGGAAACATCGAGTTCATGGGGCGTGAGGACAACCAAATCAAGCTTCGCGGATACC 2797
Db     161851 CGGACGGCTGCATCGAATTCCTCGGGCGCGTCGACAGCCAGGTCAAGCTCGGCGGCCACC 161910

Qy       2798 GCGTTGAGCTCGGGGAAATCGAGGAAACGCTCAAGTCGCATCCGAACGTACGCGACGCGG 2857
Db     161911 GCATCGAGCTGGGCGAGATCGAAGCCGCGCTGACGCAGCACGACGGCATCCAGGCCGCGG 161970

Qy       2858 TGATTGTGCCCGTCGGGAACGACGCGGCGAACAAGCTCCTTCTAGCCTATGTGGTCCCGG 2917
Db     161971 CCGCCACGGTCTGCGCCACAGCGGGCGGGGGACAGCAGCTGGCCGCCTTCGTCGTGCCGC 162030

Qy       2918 AGGGCACACGGAGACGCGCTGCCGAGCAGGACGCGAGCCTCAAGACCGAGCGGATCGACG 2977
Db     162031 GACACGCCCCG---GCTTCTGCCGACGGTGCCGCGCACGATGCCGCCTGGCAGGCCCTGG 162087

Qy       2978 CGAGAGCACACGCCGCCGAAGCGGACGGCTTGAGCGACGGCGAGAGGGTG 3027
Db     162088 CGGCGGCCTGCGCGTCCGCGGCGGCGGCGCTGCCCGCGCCGGCCAAGGTG 162137
```

Search completed: October 4, 2003, 03:30:57
Job time : 15178.8 secs

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:01 ; Search time 2488.11 Seconds
(without alignments)
12307.996 Million cell updates/sec

Title: US-10-014-717-1_COPY_62369_63628
Perfect score: 1260
Sequence: 1 atgacacaggagcaagcgaa..........agccctccaaagctggatag 1260

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	115	9.1	1285	29	BZ567880	BZ567880 pacs2-164
	2	84.6	6.7	975	29	BZ675691	BZ675691 PUBDH78TD
	3	63	5.0	330	28	AY174775	AY174775 AY174775
	4	55.4	4.4	1127	29	BZ548452	BZ548452 pacs1-60_
	5	54.8	4.3	925	29	CNS0091P	AL053013 Drosophil
c	6	54.2	4.3	1258	29	BZ573154	BZ573154 msh2_2968
	7	52	4.1	609	12	BI716406	BI716406 1031010A1
c	8	51.8	4.1	925	29	CNS0091P	AL053013 Drosophil
	9	50.8	4.0	425	10	BG488189	BG488189 RHIZ2_60_
c	10	50.6	4.0	1170	29	BZ569411	BZ569411 pacs2-164
	11	50.2	4.0	512	14	CD045174	CD045174 psHB018xC
	12	50.2	4.0	512	14	CD045351	CD045351 psHB019xF
	13	50.2	4.0	596	14	CD260296	CD260296 psMA008xI
	14	50.2	4.0	610	14	CD043612	CD043612 psHB009xH
	15	50.2	4.0	623	14	CD262676	CD262676 psMA018xK
	16	50.2	4.0	715	14	CD041124	CD041124 psHB001xJ
c	17	49.8	4.0	1242	29	BZ579496	BZ579496 msh2_6381
	18	49.6	3.9	638	14	CD486072	CD486072 CFUS5.4D0
c	19	49.6	3.9	691	10	BE559050	BE559050 HV_CEb002
	20	49.4	3.9	932	29	CNS0072Q	AL066742 Drosophil
	21	49.2	3.9	402	10	BG559249	BG559249 RHIZ2_52_
	22	49	3.9	665	12	BM370205	BM370205 EBro08_SQ
	23	48.8	3.9	788	14	CB668866	CB668866 OSJNEd16O
	24	48.4	3.8	899	13	BX360973	BX360973 BX360973
	25	48	3.8	641	9	AU092915	AU092915 AU092915
c	26	47.8	3.8	572	14	CA404475	CA404475 EL01N0518
c	27	47.8	3.8	583	10	BE761796	BE761796 947006H01
	28	47.8	3.8	799	29	BZ977578	BZ977578 PUGIS17TD
c	29	47.8	3.8	838	29	BZ977576	BZ977576 PUGIS17TB
	30	47.2	3.7	389	10	BE229868	BE229868 99AS102 R
	31	47.2	3.7	406	9	AU063730	AU063730 AU063730
	32	47.2	3.7	413	9	AU093702	AU093702 AU093702
	33	47.2	3.7	417	14	D22231	D22231 RICC10540A
	34	47.2	3.7	421	9	AU069623	AU069623 AU069623
	35	47.2	3.7	442	9	AU182095	AU182095 AU182095
	36	47.2	3.7	442	9	AU182164	AU182164 AU182164
	37	47.2	3.7	446	9	AU182068	AU182068 AU182068
	38	47.2	3.7	449	9	AU161375	AU161375 AU161375
	39	47.2	3.7	549	10	BF728783	BF728783 1000067D0
	40	47.2	3.7	643	9	AW585283	AW585283 707094C03
	41	47.2	3.7	707	9	AU161432	AU161432 AU161432
	42	47.2	3.7	719	12	BM600174	BM600174 170006870
c	43	47.2	3.7	738	14	CB635178	CB635178 OSIIEb15E
	44	47.2	3.7	741	14	CD428040	CD428040 ETH1_32_D
c	45	47.2	3.7	764	14	CB668867	CB668867 OSJNEd16O

ALIGNMENTS

RESULT 1
BZ567880
LOCUS BZ567880 1285 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_7256.x1 pacs2-164 Pseudomonas aeruginosa genomic clone pacs2-164_7256, genomic survey sequence.
ACCESSION BZ567880
VERSION BZ567880.1 GI:27199936
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1285)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers

```
    source          1..1285
                    /organism="Pseudomonas aeruginosa"
                    /mol_type="genomic DNA"
                    /strain="2-164"
                    /db_xref="taxon:287"
                    /clone="pacs2-164_7256"
                    /clone_lib="pacs2-164"
                    /note="clinical isolate 2-164 Whole genomic shotgun
                    library."
BASE COUNT      219 a    405 c    355 g    289 t     17 others
ORIGIN

  Query Match             9.1%;  Score 115;  DB 29;  Length 1285;
  Best Local Similarity   53.3%;  Pred. No. 1.6e-16;
  Matches  264;  Conservative    0;  Mismatches  230;  Indels    1;  Gaps    1;

Qy        670 GACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAG 729
Db         95 GACGTCTACAGCGGGCTGGTGCAGGCCGCCGACGAGAGCGGCCAGTTGAGCGAAGCGGAA 154

Qy        730 CTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATC 789
Db        155 CTCGTCTCCATGGCCCACCTGCTGATGATGAGCGGCTTCGAGACCACCATGAACATGATC 214

Qy        790 GCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAG 849
Db        215 GGCAACGCGCTGGTCACCCTGCTGGTCAACCCGGAGCAACTGGCGTTGCTGCGGGCGCAG 274

Qy        850 CCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAATA 909
Db        275 CCGGAACTCCTGCCCAACGCCATGGAAGAACTGGTCCGCCACGACAGTCCGGTGCGCGCC 334

Qy        910 GGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGAAAGGG 969
Db        335 TCGATGTTGCGCTTCACCGTGGAAGACGTGGAACTGGACGGGGTCACCATTCCCGCCGGC 394

Qy        970 GAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTCTCCAGGCCA 1029
Db        395 GAATACATCCTGGTCTCCAACCTGACCGCCAACCACGATGCCGAGCGCTTCGACGATCCC 454

Qy       1030 GACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCCCCAT 1089
Db        455 GACCGCCTCGACCTCACCCGCAACACCGATGGCCATCTCGGCTACGGCTTCGGCGTGCAC 514

Qy       1090 GTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCG-GAGATCGCCGTGGGCACCATCTT 1148
Db        515 TACTGCGTCGGCGCCTCGCTGGCCCGGCTGGAGGGGCGGGATCGCCATCCAGCGCCTGCT 574

Qy       1149 CCGTAGGTTCCCCGA 1163
Db        575 CGCGCGCTTCCCCGA 589

RESULT 2
BZ675691
LOCUS       BZ675691                 975 bp    DNA     linear   GSS 05-FEB-2003
DEFINITION  PUBDH78TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa029M12,
            genomic survey sequence.
ACCESSION   BZ675691
VERSION     BZ675691.1  GI:28226790
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 975)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
            ,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..975
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa029M12"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
BASE COUNT      173 a    323 c    302 g    177 t
ORIGIN

  Query Match             6.7%;  Score 84.6;  DB 29;  Length 975;
  Best Local Similarity   49.3%;  Pred. No. 1.9e-09;
  Matches  250;  Conservative    0;  Mismatches  254;  Indels    3;  Gaps    1;

Qy        668 ATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGG 727
Db        216 ATGACTTCATGACCGCCCTGACCCAAGCCGACGAAAAAGGCGAGTCCCTCAACGAGGAGC 275

Qy        728 AGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTA 787
Db        276 AATTGCTGCTGTTGTGCATTGCGATCCTGGTCGCCGGTTACGAAGGCAGTGCCTCGCAGA 335

Qy        788 TCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCG 847
Db        336 TCCCCAACTTTATCCAGGTGCTGCTGGACAACCCCGCGCAGTGGCAACAGCTCAAGGCCG 395

Qy        848 AGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAA 907
Db        396 ATCCCGAGCAGATACCGCAGGCGGTCGAAGAGCTGCTGCGCTACATCCCGCTGGCCTCGG 455

Qy        908 TAGGAA---CTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGA 964
Db        456 CGGCGATGTTTGTGCACTACGCGCTGGAAGACATTCAGGTCGGTGAAACCCTGGTGCGCC 515

Qy        965 AAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTCTCCA 1024
Db        516 AGGGCGACGCGGTGTTCGCCTCGATCGGCGCCGCCAACCATGACCCGGCCCGCTTTGAAA 575

Qy       1025 GGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCC 1084
Db        576 ACCCACAGGCGCTGGACCTGCAACGGGATGCCAGCGGCCATTTCGGCTTTGGCCACGGCC 635

Qy       1085 CCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCACCA 1144
Db        636 TGCATCACTGCATCGGCTCAGCCCTGGCGCGGGTCGAATTGCAGGAGGCGCTGCACGCCC 695

Qy       1145 TCTTCCGTAGGTTCCCCGAGATGAAGC 1171
Db        696 TGGTGGTGCGCCTGCCCAGCCTGCAGC 722

RESULT 3
AY174775
LOCUS       AY174775                 330 bp    DNA     linear   GSS 13-JAN-2003
DEFINITION  AY174775 Mycobacterium avium subsp. paratuberculosis DNA
            Mycobacterium avium subsp. paratuberculosis genomic clone Mptb61.15
            , genomic survey sequence.
ACCESSION   AY174775
VERSION     AY174775.1  GI:27689314
KEYWORDS    GSS.
SOURCE      Mycobacterium avium subsp. paratuberculosis
  ORGANISM  Mycobacterium avium subsp. paratuberculosis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
```

```
             avium complex (MAC).
REFERENCE    1  (bases 1 to 330)
  AUTHORS    Nielsen,K.K. and Ahrens,P.
  TITLE      Subtractive hybridization and bioinformatics: combining genetic
             methods to obtain new information on species-specific elements of
             Mycobacterium avium subspecies paratuberculosis
  JOURNAL    Unpublished
COMMENT      Contact: Nielsen KK
             Bacteriology
             Danish Veterinary Institute
             Bulowsvej 27, Copenhagen V, 1790, Denmark
             Email: kni@dvi.dk
             Class: unknown.
FEATURES              Location/Qualifiers
     source           1..330
                      /organism="Mycobacterium avium subsp. paratuberculosis"
                      /mol_type="genomic DNA"
                      /sub_species="paratuberculosis"
                      /db_xref="taxon:1770"
                      /clone="Mptb61.15"
                      /clone_lib="Mycobacterium avium subsp. paratuberculosis
                      DNA"
BASE COUNT       70 a     96 c     93 g     71 t
ORIGIN

  Query Match             5.0%;  Score 63;  DB 28;  Length 330;
  Best Local Similarity   52.9%;  Pred. No. 0.00016;
  Matches  135;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        637 CTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCC 696
Db         14 CTCGCCGTGCGCAAGCGCGCGGAGCCAGGTGACGATATCGCGACTTCCTTACTCCACGCC 73

Qy        697 GAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATC 756
Db         74 GAGGTTGATGGTCAGAGCCTCAGCGATTTGGAATTTAACCTGTTCTTTATGTTGCTGATC 133

Qy        757 GCTGCTGGCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGG 816
Db        134 GATGCCTGAGGTGATACGACCCGCAATCTTGTAGCAGCGGGAATTCTTGCTCTGCTTGAG 193

Qy        817 TCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGAT 876
Db        194 CATCCCCAAGAACTGCAGCGGCTTAAGGCAGACCCTTCACTCATGCCGACCGCGATCGAA 253

Qy        877 GAGGTGCTCCGCTTC 891
Db        254 GAAATGCTGCGCTAC 268


RESULT 4
BZ548452
LOCUS        BZ548452               1127 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION   pacs1-60_1124.s1 pacs1-60 Pseudomonas aeruginosa genomic clone
             pacs1-60_1124, genomic survey sequence.
ACCESSION    BZ548452
VERSION      BZ548452.1  GI:27152033
KEYWORDS     GSS.
SOURCE       Pseudomonas aeruginosa
  ORGANISM   Pseudomonas aeruginosa
             Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
             Pseudomonadaceae; Pseudomonas.
REFERENCE    1  (bases 1 to 1127)
  AUTHORS    Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
             Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE      Whole-Genome-Sequence variation among multiple isolates of
             Psedomonas aeruginosa library
  JOURNAL    J. Bacteriol.,  (2002) In press
COMMENT      Contact: Chris K. Raymond
             Genome Center
             University of Washington
             Box 352145, Seattle, WA 98105-2145, USA
             Tel: 2062216954
             Fax: 2066857244
             Email: craymond@u.washington.edu
             Class: shotgun.
FEATURES              Location/Qualifiers
     source           1..1127
                      /organism="Pseudomonas aeruginosa"
                      /mol_type="genomic DNA"
                      /strain="1-60"
                      /db_xref="taxon:287"
                      /clone="pacs1-60_1124"
                      /clone_lib="pacs1-60"
                      /note="clinical isolate 1-60 Whole genomic shotgun
                      library."
BASE COUNT      197 a    379 c    340 g    210 t      1 others
ORIGIN

  Query Match             4.4%;  Score 55.4;  DB 29;  Length 1127;
  Best Local Similarity   48.3%;  Pred. No. 0.014;
  Matches  186;  Conservative    0;  Mismatches  196;  Indels    3;  Gaps    1;

Qy        284 TGTTCGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTA 343
Db        252 TGCTCAACCTCGACCCGCCGGACCATACCCGCCTGCGCTCGCTGGTTGGCCGCGCGTTCA 311

Qy        344 CGTCACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCG 403
Db        312 CCCCGCGCCAGGTGGAGCGCCTGCAACCGCATATACAACGGATCACCGAGGAATTGCTGG 371

Qy        404 ATGCTCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGA 463
Db        372 ACGCCATGGCCGGCCGCGAACAGGCCGACCTGATGGCCGACTTCGCGATCCCGCTGACCA 431

Qy        464 TGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCT 523
Db        432 TCGCGGTGATCTTCGAGCTGCTGGGCATTCCCGAGGCCGAGCGCGAACACGCCCGCCAGT 491

Qy        524 TCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGG---TGCCCCAGGTCGATGAGG 580
Db        492 CCTGGGAGCGCCAGGCGGAACTGCTGTCGCCGGACGAGGCCCAGGCCCTGGCCGATGCGC 551

Qy        581 AGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCG 640
Db        552 AGGTCGACTACCTGCGCGTGCTGCTCGAGGCCAAGCGCCGGCAGCCGGCCGACGACGTCT 611

Qy        641 ATGAGCGGCGCAGGAACCCGCTCGA 665
Db        612 ACAGCGGGCTGTTTAAGGCCGCCGA 636


RESULT 5
CNS0091P
LOCUS        CNS0091P                925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION   Drosophila melanogaster genome survey sequence TET3 end of BAC #
             BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
             fly), genomic survey sequence.
ACCESSION    AL053013
VERSION      AL053013.1  GI:4934461
KEYWORDS     GSS.
SOURCE       Drosophila melanogaster (fruit fly)
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
             Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
             Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 925)
  AUTHORS    Genoscope.
  TITLE      Direct Submission
  JOURNAL    Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
             - Web : www.genoscope.cns.fr)
COMMENT      Determination of this BAC-end sequence was carried out as part of a
             collaboration with the Berkeley Drosophila Genome Project (BDGP).
             The BDGP is constructing a physical map of the Drosophila
             melanogaster genome using these BACs. For further information
```

```
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN

  Query Match             4.3%;  Score 54.8;  DB 29;  Length 925;
  Best Local Similarity   11.8%;  Pred. No. 0.018;
  Matches   43;  Conservative  175;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        387 CGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTA 446
               :|: : |: |: :: :    :| |:: :|:  :      : : :|   ::: : :|:: :
Db        558 SGYGKGCSSGSGBSCSCCSSCSCSSSSCSCCBCCCCCSCSSYCCSSSBSSSKCSSTSBSC 617

Qy        447 CGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGA 506
               : | : ::  |:|:  :|::::: |: :: : :|:: :  |:  ::::::  :::|:|
Db        618 SCCCSSKSVCGTSCSSSSSCSSSSSSTSSSSTSSSTSKSSSSGSSSSSSSSYTTSKSTSA 677

Qy        507 CGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCC 566
               :| | ::     : :     |:::: ::: : : ::|::: :::::      :: : :
Db        678 SGSGSWSAGGGSGSTGSTSSSSSSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSSSSSTS 737

Qy        567 CCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCT 626
               : : : |:    : ::    ::   : | |  ::: ::  :: :  :  : :|: :::: |::
Db        738 SBBSCTSTSSSSSSSYSSSTCSCCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSVGTSSS 797

Qy        627 CCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCT 686
               :   |: |  || :  |  :  :::|  :  |  :: ::          ||    | :: |::
Db        798 SDSTSTCCSCCCYMCTCCSTYBMBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSSTNGMB 857

Qy        687 GCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGG 746
               |      :::::  :::::::: ::|::: : :: :: ::: ::  ::::::  :: |:|
Db        858 GTSSACSSSSSSCSSSSVSSSSKSSASSSSVSSSGSSGVSSNSSSASKSSSSGSVSSGSG 917

Qy        747 TGCG 750
                |:|
Db        918 SGSG 921
```

```
RESULT 6
BZ573154/c
LOCUS       BZ573154                1258 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  msh2_2968.y2 msh Pseudomonas aeruginosa genomic clone msh2_2968,
            genomic survey sequence.
ACCESSION   BZ573154
VERSION     BZ573154.1  GI:27208215
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1258)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1258
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_2968"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."
BASE COUNT      192 a    368 c    370 g    328 t
ORIGIN

  Query Match             4.3%;  Score 54.2;  DB 29;  Length 1258;
  Best Local Similarity   45.7%;  Pred. No. 0.027;
  Matches  188;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy        447 CGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGA 506
               | || |  |  | |    || |   |||   || | || || | |  ||   | |
Db        531 CGTGGTGTTCACCGTGTCCGACGACAAGGGCGACGTGGTCAACGTCCTGCGCGCCGGCGC 472

Qy        507 CGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCC 566
               |||    | ||  | |   | |  || | | | |||| |||||            | |
Db        471 CGACGGCTACCTGCTCAAGGACATGGAGCCGGAGCGCCTGCTCGAGCACATCCGCCAGGC 412

Qy        567 CCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCT 626
               |    ||  || |||| || || |      | ||  | |    |   |  | || |||
Db        411 CGCCACCGGACAGATGACCCTGAGCCCGCAGTTGACCCAGATCCTCGCCCAGGCCCTGCG 352

Qy        627 CCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCT 686
                  |||| || |  | |     | |  |||           ||   | |
Db        351 GGGCGACGACCGCTCGAAGAGCCTCGACGAGCTGACCGAGCGCGAACGGCAGATCCTCCG 292

Qy        687 GCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGG 746
               |     ||| | | | || ||| || | ||| | ||      |||||| |      || ||
Db        291 CCAGATCGCCCACGGCTACAGCAACAAGATGATCGCGCGAAAGCTGGACATCACCGAAGG 232

Qy        747 TGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAA 806
                || | |  |     || ||             | | | ||   | |    |  | || |
Db        231 CACGGTGAAGGTCCATGTCAAGCGGGTGCTGCACAAGCTCGGCATGCGTTCGCGGGTCGA 172

Qy        807 CCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCGGGCT 857
                      |  || | | | || |  ||| |||||| | |||| || ||||||
Db        171 GGCCGCGGTGTGGGCGGTGGAGAACGACCTGGTGTAAGCCGCGCGCGGGCT 121

RESULT 7
BI716406
LOCUS       BI716406                 609 bp    mRNA    linear   EST 19-SEP-2001
DEFINITION  1031010A12.y1 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BI716406
VERSION     BI716406.1  GI:15692101
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 609)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished
```

```
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1..609
                     /organism="Chlamydomonas reinhardtii"
                     /mol_type="mRNA"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                     ), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; Stress condition II library, constructed by John
                     Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                     cells grown to mid-log phase in TAP (NH4+ - containing)
                     and shifted to TAP - NO3- (24hrs); H2 production
                     conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                     Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                     sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                     sites. pBluescript II SK- plasmids were excised from the
                     lambda ZAP clones by superinfection with ExAssist
                     (Stratagene) phage. The library was normalized using
                     method 4 described in Bonaldo et al., (1996) Genome
                     Research 6: 791-806."
BASE COUNT      119 a    200 c    213 g     77 t
ORIGIN

  Query Match             4.1%;  Score 52;  DB 12;  Length 609;
  Best Local Similarity   48.9%;  Pred. No. 0.073;
  Matches  139;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        568 CAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTC 627
Db         32 CAGCTCAGCGGCGGGTGGCGCATGCGGGTGGCACTGGGCTGCGCGCTGCTGGCCCGGCCG 91

Qy        628 CATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTG 687
Db         92 CACCTGCTGCTGCTGGACGAGCCCACCAACCACCTGGACATTCTGTCCATCCTGTGGCTG 151

Qy        688 CTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGT 747
Db        152 CAGAACCGCCTGGCCACTTGCTGCGAGGGGCAGACGCTGGTGCTGGTCAGCCACGACCGG 211

Qy        748 GCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAAC 807
Db        212 GCGTTCCTCAACGCGGTGACGCAGGAGACCATCATCCTGCGCGACGGCGCCCTGACCTAC 271

Qy        808 CTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCC 851
Db        272 TTCCCCGGCGCCTACGACGCCTACATGCAGGCGAAGGCGGAGGC 315

RESULT 8
CNS0091P/c
LOCUS       CNS0091P                 925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN

  Query Match             4.1%;  Score 51.8;  DB 29;  Length 925;
  Best Local Similarity   14.5%;  Pred. No. 0.092;
  Matches   52;  Conservative  159;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        530 CGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGA 589
Db        925 CSBSCSCSGSCSSBSCSSSSMSTSSSNSSBCSSCSSSSBSSSSTSSMSSSSBSSSSGSSSS 866

Qy        590 CCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGC 649
Db        865 SSGTSSACVKCNASSSCGCCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGS 806

Qy        650 GCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCA 709
Db        805 GGASASHSSSSACBSSSSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSS 746

Qy        710 GCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCG 769
Db        745 ASAGSVVSSASSSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSASCASCSCCC 686

Qy        770 ATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGC 829
Db        685 TSWSCSCSTSASMSAARSSSSSSSSSCSSSSMSASSSASSSSASSSSSSGSSSSSGSACGB 626

Qy        830 TCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGC 888
Db        625 SMSSGGGGSGSVSASSGMSSSVSSSGGRSSGSGGGGGVGGSGSSSSGSGSSGGSGSVCSC 567

RESULT 9
BG488189
LOCUS       BG488189                 425 bp    mRNA    linear   EST 27-MAR-2001
DEFINITION  RHIZ2_60_F03.b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG488189
VERSION     BG488189.1  GI:13469424
KEYWORDS    EST.
SOURCE      Sorghum propinquum
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
```

```
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: JEN REV
            High quality sequence stop: 397
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..425
                     /organism="Sorghum propinquum"
                     /mol_type="mRNA"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT       69 a    136 c    156 g     63 t      1 others
ORIGIN

  Query Match              4.0%;  Score 50.8;  DB 10;  Length 425;
  Best Local Similarity   51.1%;  Pred. No. 0.13;
  Matches  118;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy        530 CGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCAGGTCGATGAGGAGACCAAGA 589
              || || | ||  ||  ||| | | |||| | |     ||| ||  | || |       |
Db         87 CGTCGTCAGCTAGCTAGCTAGCCATGGGGTCGACGGCGGAGGACGTGGCGGCGGTGGCGG 146

Qy        590 CCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGC 649
               | ||  |||| | |    || |  ||     | |          |||||       | ||
Db        147 ACGAGGAGGCGTGCATGTACGCGATGCAGCTGGCGTCGTCGTCGATCCTCCCCATGACGC 206

Qy        650 GCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCA 709
                 | |||| |||| ||      | | |  ||  | |||| |   ||| ||  |||| ||||
Db        207 TGAAGAACGCGCTGGAGCTGGGCCTGCTGGAGGTGCTTCAGAAGGACGCCGCCGCCGGCG 266

Qy        710 GCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTG 760
              | | |   |    || ||||  ||| | |  || || |||    | ||| |
Db        267 GTAAGGCGCTGGCGGCGGAGGAGGTGGTGGNCGCGGCTGCCCGTGGCGCCG 317


RESULT 10
BZ569411/c
LOCUS       BZ569411                1170 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_8374.x1u1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_8374, genomic survey sequence.
ACCESSION   BZ569411
VERSION     BZ569411.1  GI:27203996
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1170)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
```

```
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..1170
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_8374"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT       92 a    489 c    543 g      9 t     37 others
ORIGIN

  Query Match              4.0%;  Score 50.6;  DB 29;  Length 1170;
  Best Local Similarity   44.3%;  Pred. No. 0.19;
  Matches  179;  Conservative    0;  Mismatches  225;  Indels    0;  Gaps    0;

Qy        453 GGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAA 512
              |||   |||     ||    |   | | || ||| |          |  |||||| | |  |
Db        752 GGGCGCCCCCCCCCGGCGGCCCCGCCCTCCGCGCGNCGCGTCGGCCGCGCGCCGCCCCGN 693

Qy        513 GTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGT 572
                 |||    |    || ||      |    |   |   |  | || |  |     |||  ||
Db        692 CGCCCGGGNCCCGCGCGCGCGCTCGCGGNCCCTGGCGGCTGTCGCTGCGCCCCCGGGCT 633

Qy        573 CGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGA 632
               |    | || |       | | | | ||    |    |       || |||| ||||| |  ||
Db        632 CTGGGCGGCGGNNCCGGCGCGGGGGCCCCCGCGGCCCCGCGGGCGCGCGCGGGCCCTCCG 573

Qy        633 CGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCA 692
               || || ||   |  | |   || |     |||| ||        |       | ||  || ||  |
Db        572 CGCCCGCGGGGCGGGCGGCGGCCGGCCGCCCGCGTCCCCCTCGCGGCCGCCGCGGCCCCC 513

Qy        693 GGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGAT 752
               | ||| || |  |||| ||           || ||     |  || |||||  |  |   ||
Db        512 GCCCGCGGNCCGCGGCGGCGCCGGCGGCGCGCGCCGGGCGGCCGCGCGCCCGCGGCCGGG 453

Qy        753 TATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCT 812
                 || |      | |  ||            |    ||||    ||  ||  ||   || |
Db        452 GCGCGGGGGGCCGCGCGCTCGCGCGCCNCGNCCGGGCGCGCCTGCGCTGGGCGGCCNNCG 393

Qy        813 GCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCGGGC 856
               ||    | | | |   || || ||   |  |   || |   ||||||||
Db        392 GCCCCGGGCTGCGCGGCGCGCGCGCGGGCGGGGCCGGCCCGGGC 349


RESULT 11
CD045174
LOCUS       CD045174                 512 bp    mRNA    linear   EST 09-MAY-2003
DEFINITION  psHB018xC17f_183952 psHB: Infected hypocotyl soybean host. 48 hrs
            post infection Phytophthora sojae cDNA clone sHB018C17 5, mRNA
            sequence.
ACCESSION   CD045174
VERSION     CD045174.1  GI:30498767
KEYWORDS    EST.
SOURCE      Phytophthora sojae
  ORGANISM  Phytophthora sojae
            Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
            Phytophthora.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
  TITLE     USDA-IFAFS: Expression of Phytophthora sojae genes during infection
            and propagation
  JOURNAL   Unpublished
COMMENT     Contact: Tyler B
            Tyler lab
            VBI
            1880 Pratt Dr., Blacksburg, VA 24061, USA
            Tel: 540-231-7318
            Email: bmtyler@vt.edu
```

```
             PCR PRimers
             FORWARD: BK reverse
             Plate: 018  row: C  column: 17
             Seq primer: BK reverse
             High quality sequence stop: 512.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Phytophthora sojae"
                     /mol_type="mRNA"
                     /db_xref="taxon:67593"
                     /clone="sHB018C17"
                     /tissue_type="infected host tissue"
                     /cell_line="P6497"
                     /dev_stage="48 hour post infection"
                     /clone_lib="psHB: Infected hypocotyl soybean host. 48 hrs
                     post infection"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI;
                     USDA-IFAFS:Expression of Phytophthora sojae genes during
                     infection and propagation."
BASE COUNT      105 a    159 c    182 g     66 t
ORIGIN

  Query Match             4.0%;  Score 50.2;  DB 14;  Length 512;
  Best Local Similarity   45.5%;  Pred. No. 0.18;
  Matches  178;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy        546 GCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGT 605

Db         66 GCGCGTGGTGGACTACGCCGTCAAGATCCGCGTGGAGCCCAAGGGCGTGGACCTCAAGAA 125

Qy        606 CACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGA 665

Db        126 CGTCAAGATGAGCATGAACCCCTTCTGCGAGATCGCCGTGGAGGAGGCCATCCGCCTCAA 185

Qy        666 AAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAA 725

Db        186 GGAGAAGAAGCTGGCCACCGAGGTCGTGGCCGTGTCCATCGGCCCCAAGCAGAGCCAGGA 245

Qy        726 GGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCT 785

Db        246 GACGCTGCGCACGGCGCTGGCCATGGGCGCGGACCGCGGCATCCACATCACCACGGACAT 305

Qy        786 TATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGC 845

Db        306 GCGCACGGACCAGGAGCTGCAGCCGCTGGCCGTGGCCAAGCTGCTCAAGGAGGTCGTGGC 365

Qy        846 CGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAG 905

Db        366 CAAGGAGGAGCCGCAGCTCGTCATCTGCGGCAAGCAGAGCATCGACGCCGACGCCGCGCA 425

Qy        906 AATAGGAACTGTGCGTTTCGCCAGGCAGGAC 936

Db        426 GACGGGCCCCATGCTGGCCGGCCTGCTGGAC 456


RESULT 12
CD045351
LOCUS       CD045351                 512 bp    mRNA    linear   EST 09-MAY-2003
DEFINITION  psHB019xF20f_184338 psHB: Infected hypocotyl soybean host. 48 hrs
            post infection Phytophthora sojae cDNA clone sHB019F20 5, mRNA
            sequence.
ACCESSION   CD045351
VERSION     CD045351.1  GI:30498944
KEYWORDS    EST.
SOURCE      Phytophthora sojae
  ORGANISM  Phytophthora sojae
            Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
            Phytophthora.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
  TITLE     USDA-IFAFS: Expression of Phytophthora sojae genes during infection
            and propagation
  JOURNAL   Unpublished
COMMENT     Contact: Tyler B
            Tyler lab
            VBI
            1880 Pratt Dr., Blacksburg, VA 24061, USA
            Tel: 540-231-7318
            Email: bmtyler@vt.edu
            PCR PRimers
            FORWARD: BK reverse
            Plate: 019  row: F  column: 20
            Seq primer: BK reverse
            High quality sequence stop: 512.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Phytophthora sojae"
                     /mol_type="mRNA"
                     /db_xref="taxon:67593"
                     /clone="sHB019F20"
                     /tissue_type="infected host tissue"
                     /cell_line="P6497"
                     /dev_stage="48 hour post infection"
                     /clone_lib="psHB: Infected hypocotyl soybean host. 48 hrs
                     post infection"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI;
                     USDA-IFAFS:Expression of Phytophthora sojae genes during
                     infection and propagation."
BASE COUNT      105 a    159 c    181 g     67 t
ORIGIN

  Query Match             4.0%;  Score 50.2;  DB 14;  Length 512;
  Best Local Similarity   45.5%;  Pred. No. 0.18;
  Matches  178;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy        546 GCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGT 605

Db         66 GCGCGTGGTGGACTACGCCGTCAAGATCCGCGTGGAGCCCAAGGGCGTGGACCTCAAGAA 125

Qy        606 CACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGA 665

Db        126 CGTCAAGATGAGCATGAACCCCTTCTGCGAGATCGCCGTGGAGGAGGCCATCCGCCTCAA 185

Qy        666 AAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAA 725

Db        186 GGAGAAGAAGCTGGCCACCGAGGTCGTGGCCGTGTCCATCGGCCCCAAGCAGAGCCAGGA 245

Qy        726 GGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCT 785

Db        246 GACGCTGCGCACGGCGCTGGCCATGGGCGCGGACCGCGGCATCCACATCACCACGGACAT 305

Qy        786 TATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGC 845

Db        306 GCGCACGGACCAGGAGCTGCAGCCGCTGGCCGTGGCCAAGCTGCTCAAGGAGGTCGTGGC 365

Qy        846 CGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAG 905

Db        366 CAAGGAGGAGCCGCAGCTCGTCATCTGCGGCAAGCAGAGCATCGACGCCGACGCCGCGCA 425

Qy        906 AATAGGAACTGTGCGTTTCGCCAGGCAGGAC 936

Db        426 GACGGGCCCCATGCTGGCCGGCCTGCTGGAC 456


RESULT 13
CD260296
LOCUS       CD260296                 596 bp    mRNA    linear   EST 23-MAY-2003
DEFINITION  psMA008xI18f_187110 sMA: Phytophtora sojae grown in synthetic
            medium Phytophthora sojae cDNA clone sMA008I18 5, mRNA sequence.
ACCESSION   CD260296
VERSION     CD260296.1  GI:31048123
KEYWORDS    EST.
SOURCE      Phytophthora sojae
  ORGANISM  Phytophthora sojae
            Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
            Phytophthora.
```

```
REFERENCE   1  (bases 1 to 596)
  AUTHORS   Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
  TITLE     USDA-IFAFS: Expression of Phytophthora sojae genes during infection
            and propagation
  JOURNAL   Unpublished
COMMENT     Contact: Tyler B
            Tyler lab
            VBI
            1880 Pratt Dr., Blacksburg, VA 24061, USA
            Tel: 540-231-7318
            Email: bmtyler@vt.edu
            PCR PRimers
            FORWARD: BK reverse
            Plate: 008  row: I  column: 18
            Seq primer: BK reverse
            High quality sequence stop: 596.
FEATURES             Location/Qualifiers
     source          1..596
                     /organism="Phytophthora sojae"
                     /mol_type="mRNA"
                     /db_xref="taxon:67593"
                     /clone="sMA008I18"
                     /tissue_type="mycelium"
                     /cell_line="P6497"
                     /dev_stage="mycelium"
                     /clone_lib="sMA: Phytophtora sojae grown in synthetic
                     medium"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      119 a     190 c     212 g      75 t
ORIGIN

  Query Match             4.0%;  Score 50.2;  DB 14;  Length 596;
  Best Local Similarity   45.5%;  Pred. No. 0.19;
  Matches  178;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy        546 GCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGT 605

Db         67 GCGCGTGGTGGACTACGCCGTCAAGATCCGCGTGGAGCCCAAGGGCGTGGACCTCAAGAA 126

Qy        606 CACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGA 665

Db        127 CGTCAAGATGAGCATGAACCCCTTCTGCGAGATCGCCGTGGAGGAGGCCATCCGCCTCAA 186

Qy        666 AAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAA 725

Db        187 GGAGAAGAAGCTGGCCACCGAGGTCGTGGCCGTGTCCATCGGCCCCAAGCAGAGCCAGGA 246

Qy        726 GGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCT 785

Db        247 GACGCTGCGCACGGCGCTGGCCATGGGCGCGGACCGCGGCATCCACATCACCACGGACAT 306

Qy        786 TATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGC 845

Db        307 GCGCACGGACCAGGAGCTGCAGCCGCTGGCCGTGGCCAAGCTGCTCAAGGAGGTCGTGGC 366

Qy        846 CGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAG 905

Db        367 CAAGGAGGAGCCGCAGCTCGTCATCTGCGGCAAGCAGAGCATCGACGCCGACGCCGCGCA 426

Qy        906 AATAGGAACTGTGCGTTTCGCCAGGCAGGAC 936

Db        427 GACGGGCCCCATGCTGGCCGGCCTGCTGGAC 457


RESULT 14
CD043612
LOCUS       CD043612                 610 bp    mRNA    linear   EST 09-MAY-2003
DEFINITION  psHB009xH01f_180418 psHB: Infected hypocotyl soybean host. 48 hrs
            post infection Phytophthora sojae cDNA clone sHB009H01 5, mRNA
            sequence.
ACCESSION   CD043612
VERSION     CD043612.1  GI:30497205
KEYWORDS    EST.
SOURCE      Phytophthora sojae
  ORGANISM  Phytophthora sojae
            Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
            Phytophthora.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
  TITLE     USDA-IFAFS: Expression of Phytophthora sojae genes during infection
            and propagation
  JOURNAL   Unpublished
COMMENT     Contact: Tyler B
            Tyler lab
            VBI
            1880 Pratt Dr., Blacksburg, VA 24061, USA
            Tel: 540-231-7318
            Email: bmtyler@vt.edu
            PCR PRimers
            FORWARD: BK reverse
            Plate: 009  row: H  column: 01
            Seq primer: BK reverse
            High quality sequence stop: 610.
FEATURES             Location/Qualifiers
     source          1..610
                     /organism="Phytophthora sojae"
                     /mol_type="mRNA"
                     /db_xref="taxon:67593"
                     /clone="sHB009H01"
                     /tissue_type="infected host tissue"
                     /cell_line="P6497"
                     /dev_stage="48 hour post infection"
                     /clone_lib="psHB: Infected hypocotyl soybean host. 48 hrs
                     post infection"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI;
                     USDA-IFAFS:Expression of Phytophthora sojae genes during
                     infection and propagation."
BASE COUNT      121 a     199 c     215 g      75 t
ORIGIN

  Query Match             4.0%;  Score 50.2;  DB 14;  Length 610;
  Best Local Similarity   45.5%;  Pred. No. 0.19;
  Matches  178;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy        546 GCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGT 605

Db         69 GCGCGTGGTGGACTACGCCGTCAAGATCCGCGTGGAGCCCAAGGGCGTGGACCTCAAGAA 128

Qy        606 CACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGA 665

Db        129 CGTCAAGATGAGCATGAACCCCTTCTGCGAGATCGCCGTGGAGGAGGCCATCCGCCTCAA 188

Qy        666 AAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAA 725

Db        189 GGAGAAGAAGCTGGCCACCGAGGTCGTGGCCGTGTCCATCGGCCCCAAGCAGAGCCAGGA 248

Qy        726 GGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCT 785

Db        249 GACGCTGCGCACGGCGCTGGCCATGGGCGCGGACCGCGGCATCCACATCACCACGGACAT 308

Qy        786 TATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGC 845

Db        309 GCGCACGGACCAGGAGCTGCAGCCGCTGGCCGTGGCCAAGCTGCTCAAGGAGGTCGTGGC 368

Qy        846 CGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAG 905

Db        369 CAAGGAGGAGCCGCAGCTCGTCATCTGCGGCAAGCAGAGCATCGACGCCGACGCCGCGCA 428

Qy        906 AATAGGAACTGTGCGTTTCGCCAGGCAGGAC 936

Db        429 GACGGGCCCCATGCTGGCCGGCCTGCTGGAC 459


RESULT 15
CD262676
LOCUS       CD262676                 623 bp    mRNA    linear   EST 23-MAY-2003
```

```
DEFINITION  psMA018xK21f_191994 sMA: Phytophtora sojae grown in synthetic
            medium Phytophthora sojae cDNA clone sMA018K21 5, mRNA sequence.
ACCESSION   CD262676
VERSION     CD262676.1  GI:31050503
KEYWORDS    EST.
SOURCE      Phytophthora sojae
  ORGANISM  Phytophthora sojae
            Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
            Phytophthora.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
  TITLE     USDA-IFAFS: Expression of Phytophthora sojae genes during infection
            and propagation
  JOURNAL   Unpublished
COMMENT     Contact: Tyler B
            Tyler lab
            VBI
            1880 Pratt Dr., Blacksburg, VA 24061, USA
            Tel: 540-231-7318
            Email: bmtyler@vt.edu
            PCR PRimers
            FORWARD: BK reverse
            Plate: 018  row: K  column: 21
            Seq primer: BK reverse
            High quality sequence stop: 623.
FEATURES             Location/Qualifiers
     source          1..623
                     /organism="Phytophthora sojae"
                     /mol_type="mRNA"
                     /db_xref="taxon:67593"
                     /clone="sMA018K21"
                     /tissue_type="mycelium"
                     /cell_line="P6497"
                     /dev_stage="mycelium"
                     /clone_lib="sMA: Phytophtora sojae grown in synthetic
                     medium"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      123 a    206 c    217 g     77 t
ORIGIN

  Query Match             4.0%;  Score 50.2;  DB 14;  Length 623;
  Best Local Similarity   45.5%;  Pred. No. 0.19;
  Matches  178;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy        546 GCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGT 605
              | || ||  ||||  |  |     | || ||   | |||| |||||  | |||  |
Db         59 GCGCGTGGTGGACTACGCCGTCAAGATCCGCGTGGAGCCCAAGGGCGTGGACCTCAAGAA 118

Qy        606 CACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGA 665
              |  | ||  |   |   |    | |      ||    |||   | |  |       | ||  ||| |
Db        119 CGTCAAGATGAGCATGAACCCCTTCTGCGAGATCGCCGTGGAGGAGGCCATCCGCCTCAA 178

Qy        666 AAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAA 725
                 |  |       || |  |      | |    ||||| | ||    ||||  || || ||||   | |
Db        179 GGAGAAGAAGCTGGCCACCGAGGTCGTGGCCGTGTCCATCGGCCCCAAGCAGAGCCAGGA 238

Qy        726 GGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCT 785
              |  ||||  | ||     |||      |       |        |||  |  | | |||  |    || |
Db        239 GACGCTGCGCACGGCGCTGGCCATGGGCGCGGACCGCGGCATCCACATCACCACGGACAT 298

Qy        786 TATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGC 845
                 | ||  |    | |||   | |  ||||       | || ||   |||| ||   |||     |||
Db        299 GCGCACGGACCAGGAGCTGCAGCCGCTGGCCGTGGCCAAGCTGCTCAAGGAGGTCGTGGC 358

Qy        846 CGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAG 905
              | ||   | ||     |       |        ||   ||   |   |   |   || |  |     |
Db        359 CAAGGAGGAGCCGCAGCTCGTCATCTGCGGCAAGCAGAGCATCGACGCCGACGCCGCGCA 418

Qy        906 AATAGGAACTGTGCGTTTCGCCAGGCAGGAC 936
                |  ||  |   |||      || |   || ||||
Db        419 GACGGGCCCCATGCTGGCCGGCCTGCTGGAC 449
```

Search completed: October 3, 2003, 17:48:11
Job time : 2496.11 secs

GenCore version 5.1.6


```
OM nucleic - nucleic search, using sw model

Run on:         October  3, 2003, 17:48:22 ; Search time 214.454 Seconds
                                             (without alignments)
                                             8152.990 Million cell updates/sec

Title:          US-10-014-717-1_COPY_62369_63628
Perfect score:  1260
Sequence:       1 atgacacaggagcaagcgaa..........agccctccaaagctggatag 1260

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       1770976 seqs, 693827017 residues

Total number of hits satisfying chosen parameters:      3541952

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Pending_Patents_NA_New:*
                 1:  /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
                 2:  /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
                 3:  /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
                 4:  /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
                 5:  /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
                 6:  /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
                 7:  /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	143.8	11.4	1257	6	US-10-389-647-206	Sequence 206, App
c	2	130.2	10.3	86941	1	PCT-US03-18787-2	Sequence 2, Appli
c	3	130.2	10.3	86941	1	PCT-US03-19069-2	Sequence 2, Appli
c	4	124.6	9.9	85692	1	PCT-US03-18787-1	Sequence 1, Appli
c	5	124.6	9.9	85692	1	PCT-US03-19069-1	Sequence 1, Appli
c	6	68.6	5.4	2731748	6	US-10-297-465B-1	Sequence 1, Appli
	7	55.8	4.4	2731748	6	US-10-297-465B-1	Sequence 1, Appli
	8	54	4.3	11505	7	US-60-501-821-38	Sequence 38, Appl
	9	54	4.3	73599	7	US-60-501-821-24	Sequence 24, Appl
	10	49.4	3.9	972	6	US-10-084-846A-77	Sequence 77, Appl
	11	49.4	3.9	59816	6	US-10-084-846A-1	Sequence 1, Appli
c	12	49.4	3.9	59816	6	US-10-084-846A-2	Sequence 2, Appli
	13	48.8	3.9	9369	1	PCT-US03-11231-190	Sequence 190, App
	14	48.8	3.9	9369	1	PCT-US03-11231-247	Sequence 247, App
	15	48	3.8	1036	6	US-10-425-114A-14511	Sequence 14511, A
	16	47.2	3.7	486	7	US-60-498-183-145	Sequence 145, App
	17	46.4	3.7	1215	6	US-10-425-114A-18849	Sequence 18849, A
	18	46.2	3.7	465	7	US-60-498-183-7	Sequence 7, Appli
	19	45.6	3.6	1574	6	US-10-425-114A-15634	Sequence 15634, A
	20	45	3.6	1308	6	US-10-425-114A-32476	Sequence 32476, A
	21	45	3.6	2526	6	US-10-425-114A-31445	Sequence 31445, A
	22	44.2	3.5	2181	6	US-10-425-114A-27737	Sequence 27737, A
	23	43.8	3.5	916	6	US-10-425-114A-14455	Sequence 14455, A
	24	43.8	3.5	1827	6	US-10-425-114A-21585	Sequence 21585, A
	25	43.8	3.5	2040	6	US-10-389-647-19	Sequence 19, Appl
	26	43.2	3.4	970	6	US-10-425-114A-36030	Sequence 36030, A
c	27	43.2	3.4	1263	6	US-10-425-114A-19530	Sequence 19530, A
	28	43	3.4	727	6	US-10-425-114A-32941	Sequence 32941, A
c	29	43	3.4	2167	6	US-10-425-114A-24140	Sequence 24140, A
c	30	43	3.4	40599	7	US-60-500-337-19428	Sequence 19428, A
	31	42.6	3.4	1884	6	US-10-389-647-267	Sequence 267, App
	32	42.4	3.4	603	7	US-60-498-183-13	Sequence 13, Appl
	33	42.4	3.4	1021	6	US-10-425-114A-22457	Sequence 22457, A
	34	42.4	3.4	1099	6	US-10-425-114A-15706	Sequence 15706, A
	35	42.2	3.3	783	6	US-10-425-114A-13603	Sequence 13603, A
	36	42.2	3.3	1085	6	US-10-425-114A-13465	Sequence 13465, A
	37	42.2	3.3	1191	6	US-10-425-114A-1589	Sequence 1589, Ap
	38	42.2	3.3	2368	6	US-10-425-114A-2674	Sequence 2674, Ap
	39	42.2	3.3	14800	7	US-60-479-073-496	Sequence 496, App
	40	42	3.3	537	6	US-10-425-114A-25661	Sequence 25661, A
	41	42	3.3	1047	6	US-10-425-114A-21990	Sequence 21990, A
	42	42	3.3	1121	6	US-10-425-114A-22962	Sequence 22962, A
	43	42	3.3	1243	6	US-10-425-114A-24278	Sequence 24278, A
	44	42	3.3	1464	6	US-10-425-114A-24107	Sequence 24107, A
	45	42	3.3	1491	6	US-10-425-114A-31798	Sequence 31798, A

ALIGNMENTS

```
RESULT 1
US-10-389-647-206
; Sequence 206, Application US/10389647
; GENERAL INFORMATION:
;  APPLICANT: GREENBERG, E. Peter
;  APPLICANT:  SCHUSTER, Martin
;  APPLICANT:  LOSTROH, Candi
;  TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
;  FILE REFERENCE: UIZ-038CP
;  CURRENT APPLICATION NUMBER: US/10/389,647
;  CURRENT FILING DATE:  2003-03-14
;  PRIOR APPLICATION NUMBER: 09/653730
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: 60/153022
;  PRIOR FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 710
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 206
;   LENGTH: 1257
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-10-389-647-206

  Query Match             11.4%;  Score 143.8;  DB 6;  Length 1257;
  Best Local Similarity   48.7%;  Pred. No. 6e-29;
  Matches  455;  Conservative    0;  Mismatches  447;  Indels   33;  Gaps    1;

Qy        254 CCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCGGAGGATCACGCTC 313
              | | | || ||||       |   ||     || ||   ||     |||  ||| || | |
Db        233 CGGGGATCGGCGAGGGACTCAGCCACCACATGCTCAACCTCGACCCGCCGGACCATACCC 292

Qy        314 GGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGCGCGCCG 373
              | |||  || ||  ||     |||| || |||||  ||  || |||   || |  ||||
Db        293 GCCTGCGCTCGCTGGTTGGCCGCGCGTTCACCCCGCGCCAGGTGGAGCGCCTGCAACCGC 352

Qy        374 AAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACG 433
              | ||| | || | | |||    |||| || ||       |||| |  ||  ||  ||||
Db        353 ATATAGAACGGATCACCGAGGCATTGCTGGACGCCATGGCCGGCCGCGAACAGGCCGACC 412

Qy        434 TTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTC 493
              | ||    || | ||||       | | ||     | |||| ||  ||| ||    |||
Db        413 TGATGGCCGACTTCGCGATCCCGCTGACCATCGCGGTGATCTTCGAGCTGCTGGGCATTC 472

Qy        494 CGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCTCGGCG 553
              | |  |         |||  |    ||| |  | | |
Db        473 CCGAGGCCGAGCGCGAACACGCCCGCCAGTCCTG------------------------- 506

Qy        554 TGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGG 613
                    || || ||||| ||    |  | |    |     || |   | || |   |||| |
```

```
Db        507 -------GGAGCGCCAGGCGGAACTGCTGTCGCCGGAGGAGGCCCAGGCCCTGGCCGATG 559

Qy        614 GGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACG 673

Db        560 CGCAGGTCGACTACCTGCGCGTGCTGCTCGAGGCCAAGCGCCGGCAGCCGGCCGACGACG 619

Qy        674 TCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGG 733

Db        620 TCTACAGCGGGCTGGTGCAGGCCGCCGACGAGAGCGGCCAGTTGAGCGAAGCGGAACTCG 679

Qy        734 TCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCGT 793

Db        680 TCTCCATGGCCCACCTGCTGATGATGAGCGGCTTCGAGACCACCATGAACATGATCGGCA 739

Qy        794 TCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCG 853

Db        740 ACGCGCTGGTCACCCTGCTGGTCAACCCGGAGCAACTGGCGTTGCTGCGGGCGCAGCCGG 799

Qy        854 GGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAATAGGAA 913

Db        800 AACTCCTGCCCAACGCCATGGAAGAACTGGTCCGCCACGACAGTCCGGTGCGCGCCTCGA 859

Qy        914 CTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGGAGA 973

Db        860 TGTTGCGCTTCACCGTGGAAGACGTGGAACTGGACGGGGTCACCATTCCCGCCGGCGAAT 919

Qy        974 TGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTCTCCAGGCCAGACG 1033

Db        920 ACATCCTGGTCTCCAACCTGACCGCCAACCACGATGCCGAGCGCTTCGACGATCCCGACC 979

Qy       1034 TGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCCCCATGTCT 1093

Db        980 GCCTCGACCTCACCCGCAACACCGATGGCCATCTCGGCTACGGCTTCGGCGTGCACTACT 1039

Qy       1094 GCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCACCATCTTCCGTA 1153

Db       1040 GCGTCGGCGCCTCGCTGGCCCGGCTGGAGGGGCGGATCGCCATCCAGCGCCTGCTCGCGC 1099

Qy       1154 GGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTG 1188

Db       1100 GCTTCCCCGACCTCCAGTTGGCGGTGCCCCACGCG 1134
```

RESULT 2
PCT-US03-18787-2/c
; Sequence 2, Application PC/TUS0318787
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES
; APPLICANT: Hutchinson, Richard C.
; APPLICANT: Reid, Ralph C.
; APPLICANT: Hu, Zhihao
; APPLICANT: Rascher, Andreas
; APPLICANT: Schirmer, Andreas
; APPLICANT: McDaniel, Robert
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
; TITLE OF INVENTION: PRO-GELDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF
; FILE REFERENCE: 300622009740
; CURRENT APPLICATION NUMBER: PCT/US03/18787
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,255
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/393,929
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/395,275
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/415,326
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/420,820
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/433,130
; PRIOR FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 86941
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
PCT-US03-18787-2

```
  Query Match             10.3%;  Score 130.2;  DB 1;  Length 86941;
  Best Local Similarity   50.4%;  Pred. No. 1.1e-24;
  Matches  448;  Conservative    0;  Mismatches  408;  Indels   33;  Gaps    4;

Qy        303 GGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCT 362

Db      82492 GGAGCACGCCCGGATCCGCCGCATGCTCACCGCGAACTTCACCATCCGGCGCACCGAGGC 82433

Qy        363 GCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGA 422

Db      82432 GATGCGGCCGATGATCCAGGGCATCGTCGACGGCCTCCTGGACCGGCTGATCGCCCAGGG 82373

Qy        423 ---GGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGC 479

Db      82372 CCCGCCGGCCGACCTGGTGGCCGACTTCGCCTTCCCCCTGCCGTCCCAGGTGATCGGCGT 82313

Qy        480 TCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGC 539

Db      82312 GATGCTGGGGGTCTCGGACGCCGACTTCGCGGAGTTCCAGCAG--------------GC 82268

Qy        540 GCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGC 599

Db      82267 GTCGCAGGGCGTCATGGACTTCACCGCGTCGGCCGAGGAGATGGGCGCCGCGCTCGGCGT 82208

Qy        600 GTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCC 659

Db      82207 CATGGTGGACTACGTGACCCGGATGTGCGCGGCCAAGCGCGCCGACCCGGGCGACGATCT 82148

Qy        660 GCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAG 719

Db      82147 GCTCA------------GCCGGCTCATCGTCGACCAGGAGCCGACGG--GCGGGCTCAC 82103

Qy        720 CACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGAT 779

Db      82102 CCAGCAGCAGGTGGTCGCCACCGCCCTGGTGCTGCTGCTGGCCGGGCACGAGACCACCGC 82043

Qy        780 CTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGT 839

Db      82042 CAACATGATCGCCCTGTCCACCGTCCTGCTGCTGAACCACCCCGACCAGCTCGCCCGGCT 81983

Qy        840 GAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATAT 899

Db      81982 GCGGGCCGATGCCGGGCTGATGGGCAACGCGGTGGACGAACTGCTCCGGTACATCACGAT 81923

Qy        900 CCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGAT 959

Db      81922 CGTCCAGGAAGGCACCGGACGGGTGGCCGTCGAGGACGTCGAGGTGGGCGGCGTACTCAT 81863

Qy        960 CAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATT 1019

Db      81862 CCCGGCCGGTGAAGGGGTGATCATCAACCTGCCCAGCGCCAACCGGGACCCCCACTTCGC 81803

Qy       1020 CTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAG 1079

Db      81802 GGACGCCCACGAGCTGGACCTGAGCCGGCCCAACGCCCGCGAGCATGTCGCGTTCGGCTT 81743

Qy       1080 AGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGG 1139

Db      81742 CGGCGTGCACCAGTGCCTGGGGCAGACCCTCGCCCGGGTCGAGCTCCAGATCGCTCTGGA 81683

Qy       1140 CACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTG 1188

Db      81682 GACCCTGCTGCGCCGGCTGCCCACGCTGCGCCTGGAGGTCCCGCTCGAG 81634
```

```
RESULT 3
PCT-US03-19069-2/c
; Sequence 2, Application PC/TUS0319069
; GENERAL INFORMATION:
;  APPLICANT: KOSAN BIOSCIENCES
;  APPLICANT:  Hutchinson, Richard C.
;  APPLICANT:  Reid, Ralph C.
;  APPLICANT:  Hu, Zhihao
;  APPLICANT:  Rascher, Andreas
;  APPLICANT:  Schirmer, Andreas
;  APPLICANT:  McDaniel, Robert
;  TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
;  TITLE OF INVENTION:  PRO-GELDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND
;  TITLE OF INVENTION:  ACCESSORY PROTEINS, AND USES THEREOF
;  FILE REFERENCE: 300622009740
;  CURRENT APPLICATION NUMBER: PCT/US03/19069
;  CURRENT FILING DATE:  2003-06-16
;  PRIOR APPLICATION NUMBER: US 60/389,255
;  PRIOR FILING DATE: 2002-06-14
;  PRIOR APPLICATION NUMBER: US 60/393,929
;  PRIOR FILING DATE: 2002-07-03
;  PRIOR APPLICATION NUMBER: US 60/395,275
;  PRIOR FILING DATE: 2002-07-12
;  PRIOR APPLICATION NUMBER: US 10/212,962
;  PRIOR FILING DATE: 2002-08-05
;  PRIOR APPLICATION NUMBER: US 60/415,326
;  PRIOR FILING DATE: 2002-09-30
;  PRIOR APPLICATION NUMBER: US 60/420,820
;  PRIOR FILING DATE: 2002-10-24
;  PRIOR APPLICATION NUMBER: US 60/433,130
;  PRIOR FILING DATE: 2002-12-13
;  NUMBER OF SEQ ID NOS: 153
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 86941
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
PCT-US03-19069-2

  Query Match             10.3%;  Score 130.2;  DB 1;  Length 86941;
  Best Local Similarity   50.4%;  Pred. No. 1.1e-24;
  Matches  448;  Conservative    0;  Mismatches  408;  Indels   33;  Gaps    4;

Qy        303 GGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCT 362

Db      82492 GGAGCACGCCCGGATCCGCCGCATGCTCACCGCGAACTTCACCATCCGGCGCACCGAGGC 82433

Qy        363 GCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGA 422

Db      82432 GATGCGGCCGATGATCCAGGGCATCGTCGACGGCCTCCTGGACCGGCTGATCGCCCAGGG 82373

Qy        423 ---GGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGC 479

Db      82372 CCCGCCGGCCGACCTGGTGGCCGACTTCGCCTTCCCCCTGCCGTCCCAGGTGATCGGCGT 82313

Qy        480 TCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGC 539

Db      82312 GATGCTGGGGGTCTCGGACGCCGACTTCGCGGAGTTCCAGCAG---------------GC 82268

Qy        540 GCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGC 599

Db      82267 GTCGCAGGGCGTCATGGACTTCACCGCGTCGGCCGAGGAGATGGGCGCCGCGCTCGGCGT 82208

Qy        600 GTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCC 659

Db      82207 CATGGTGGACTACGTGACCCGGATGTGCGCGGCCAAGCGCGCCGACCCGGGCGACGATCT 82148

Qy        660 GCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAG 719

Db      82147 GCTCA-------------GCCGGCTCATCGTCGACCAGGAGCCGACGG--GCGGGCTCAC 82103

Qy        720 CACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGAT 779

Db      82102 CCAGCAGCAGGTGGTCGCCACCGCCCTGGTGCTGCTGCTGGCCGGGCACGAGACCACCGC 82043

Qy        780 CTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGT 839

Db      82042 CAACATGATCGCCCTGTCCACCGTCCTGCTGCTGAACCACCCCGACCAGCTCGCCCGGCT 81983

Qy        840 GAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATAT 899

Db      81982 GCGGGCCGATGCCGGGCTGATGGGCAACGCGGTGGACGAACTGCTCCGGTACATCACGAT 81923

Qy        900 CCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGAT 959

Db      81922 CGTCCAGGAAGGCACCGGACGGGTGGCCGTCGAGGACGTCGAGGTGGGCGGCGTACTCAT 81863

Qy        960 CAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATT 1019

Db      81862 CCCGGCCGGTGAAGGGGTGATCATCAACCTGCCCAGCGCCAACCGGGACCCCCACTTCGC 81803

Qy       1020 CTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAG 1079

Db      81802 GGACGCCCACGAGCTGGACCTGAGCCGGCCCAACGCCCGCGAGCATGTCGCGTTCGGCTT 81743

Qy       1080 AGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGG 1139

Db      81742 CGGCGTGCACCAGTGCCTGGGGCAGACCCTCGCCCGGGTCGAGCTCCAGATCGCTCTGGA 81683

Qy       1140 CACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTG 1188

Db      81682 GACCCTGCTGCGCCGGCTGCCCACGCTGCGCCTGGAGGTCCCGCTCGAG 81634

RESULT 4
PCT-US03-18787-1/c
; Sequence 1, Application PC/TUS0318787
; GENERAL INFORMATION:
;  APPLICANT: KOSAN BIOSCIENCES
;  APPLICANT:  Hutchinson, Richard C.
;  APPLICANT:  Reid, Ralph C.
;  APPLICANT:  Hu, Zhihao
;  APPLICANT:  Rascher, Andreas
;  APPLICANT:  Schirmer, Andreas
;  APPLICANT:  McDaniel, Robert
;  TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
;  TITLE OF INVENTION:  PRO-GELDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND
;  TITLE OF INVENTION:  ACCESSORY PROTEINS, AND USES THEREOF
;  FILE REFERENCE: 300622009740
;  CURRENT APPLICATION NUMBER: PCT/US03/18787
;  CURRENT FILING DATE:  2003-06-13
;  PRIOR APPLICATION NUMBER: US 60/389,255
;  PRIOR FILING DATE: 2002-06-14
;  PRIOR APPLICATION NUMBER: US 60/393,929
;  PRIOR FILING DATE: 2002-07-03
;  PRIOR APPLICATION NUMBER: US 60/395,275
;  PRIOR FILING DATE: 2002-07-12
;  PRIOR APPLICATION NUMBER: US 10/212,962
;  PRIOR FILING DATE: 2002-08-05
;  PRIOR APPLICATION NUMBER: US 60/415,326
;  PRIOR FILING DATE: 2002-09-30
;  PRIOR APPLICATION NUMBER: US 60/420,820
;  PRIOR FILING DATE: 2002-10-24
;  PRIOR APPLICATION NUMBER: US 60/433,130
;  PRIOR FILING DATE: 2002-12-13
;  NUMBER OF SEQ ID NOS: 153
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 85692
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
PCT-US03-18787-1

  Query Match              9.9%;  Score 124.6;  DB 1;  Length 85692;
  Best Local Similarity   48.0%;  Pred. No. 3.5e-23;
```

```
Matches  500;  Conservative    0;  Mismatches  514;  Indels   27;  Gaps    4;

Qy        139 GGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCGGTGTTCCGCGACGAACGC 198
Db       4166 GTCCACGCCTGGCTCGTCACCGGCTACGACGCCGCGCACTCCGCGCTGGGCGACGACCGC 4107

Qy        199 TTCGCGGTCAGTCGAGAAGAGTGGAATCGAGCGCG---GAGTACTCGTCGGCCATTCCC 255
Db       4106 CTCGGCAAGAACCACGACCGGGGCAACGACCGCTGGCGCGCCCGCGCCTCGATCATGCCC 4047

Qy        256 GAGCTCAGCGATATGAAG---AAGTACGGATTGTTCGGGCTGCCGCCGGAGGATCACGCT 312
Db       4046 GAGCCCCAGCACTCCCAGCTCCAGGTCCACCTCCTCCACCAGGACCCGCCGGACCACACC 3987

Qy        313 CGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGCGCGCC 372
Db       3986 CGCATGCGGCGCCTCGTCACCGACGCCTTCACCCCGCGCCGCGTCGAGCATCTGAGGCCC 3927

Qy        373 GAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGAC 432
Db       3926 CGCTTCCAGGAGCTGGCCGACACCCTCGTCGACGCCCTCCCGGAAACCGGCCCCGCGGAC 3867

Qy        433 GTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTT 492
Db       3866 CTGGTCACCGGCTTCGCCGCCCGCTTCCCCTTGCAGGTCCTCGCCGAAGTCATCGGCCTG 3807

Qy        493 CCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGC 552
Db       3806 CCACCGGAGTCGGCGGCCCGCTTCGACCGCGACTGGGGCAAGGTCGTCCAGCCGGTCGGA 3747

Qy        553 GTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAG 612
Db       3746 CCG-----------------ACCGACCCCGGACGCCCGCTGTACGAGGCCCGCCTGCAC 3705

Qy        613 GGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGAC 672
Db       3704 GGGCTCCAGAGCTACATAGCCGAGGTCGTCG---CCCATAAGCGCGAACACGGGGAGGAC 3648

Qy        673 GTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTG 732
Db       3647 GACCTGCTCAGCCGCCTGGTCGTGGCCCGCGACCGCAGCGAACTCTCCCAGGAGGAGCTG 3588

Qy        733 GTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCG 792
Db       3587 GACTCGATGATCTTCCAGCTCCTCGTGGCGGGCCAGGAACCGGTCACCAACCAGATCACC 3528

Qy        793 TTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCC 852
Db       3527 ACCGCCCTGATCGCCCTCTTCCGCCACCCCGCCGAACTCGCCCGGCTGCGCGACGACCCG 3468

Qy        853 GGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAATAGGA 912
Db       3467 GAGCTGATGCCCCGCGCGGTCGACGAACTCCTCCGCTACGACAGCGCCTTCGAGCTGACC 3408

Qy        913 ACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGGAG 972
Db       3407 ACCTGGCGCTTCTTCGACCAGGACAGCGACCTGCACGGCACGGACGTCCCGGCCGGGGAC 3348

Qy        973 ATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGACTGTATTCTCCAGGCCAGAC 1032
Db       3347 TCGGTGATCATCTCCCTGTGCGCCGCCAACCGCGACCCACGCCGCTTCCCCGACCCCGAC 3288

Qy       1033 GTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCCCCATGTC 1092
Db       3287 ACCCTCGACCTGGACCGCTCACCCAACCCCCACCTGGCCTTCGGTCACGGCATCCACTTC 3228

Qy       1093 TGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCACCATCTTCCGT 1152
Db       3227 TGCCCCGGCGCGGCCCTGGCCCGCGCCGAACTCCAGATCGCGCTGGGCACCCTCCTGTCC 3168

Qy       1153 AGGTTCCCCGAGATGAAGCTG 1173
Db       3167 CGCCTCCCCGGAGCGCACCTG 3147
```

RESULT 5
PCT-US03-19069-1/c
; Sequence 1, Application PC/TUS0319069
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES
; APPLICANT: Hutchinson, Richard C.
; APPLICANT: Reid, Ralph C.
; APPLICANT: Hu, Zhihao
; APPLICANT: Rascher, Andreas
; APPLICANT: Schirmer, Andreas
; APPLICANT: McDaniel, Robert .
; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
; TITLE OF INVENTION: PRO-GELDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND
; TITLE OF INVENTION: ACCESSORY PROTEINS, AND USES THEREOF
; FILE REFERENCE: 300622009740
; CURRENT APPLICATION NUMBER: PCT/US03/19069
; CURRENT FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: US 60/389,255
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/393,929
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/395,275
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/415,326
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/420,820
; PRIOR FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/433,130
; PRIOR FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 85692
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
PCT-US03-19069-1

```
Query Match             9.9%;  Score 124.6;  DB 1;  Length 85692;
Best Local Similarity   48.0%;  Pred. No. 3.5e-23;
Matches  500;  Conservative    0;  Mismatches  514;  Indels   27;  Gaps    4;

Qy        139 GGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCGGTGTTCCGCGACGAACGC 198
Db       4166 GTCCACGCCTGGCTCGTCACCGGCTACGACGCCGCGCACTCCGCGCTGGGCGACGACCGC 4107

Qy        199 TTCGCGGTCAGTCGAGAAGAGTGGAATCGAGCGCG---GAGTACTCGTCGGCCATTCCC 255
Db       4106 CTCGGCAAGAACCACGACCGGGGCAACGACCGCTGGCGCGCCCGCGCCTCGATCATGCCC 4047

Qy        256 GAGCTCAGCGATATGAAG---AAGTACGGATTGTTCGGGCTGCCGCCGGAGGATCACGCT 312
Db       4046 GAGCCCCAGCACTCCCAGCTCCAGGTCCACCTCCTCCACCAGGACCCGCCGGACCACACC 3987

Qy        313 CGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGCGCGCC 372
Db       3986 CGCATGCGGCGCCTCGTCACCGACGCCTTCACCCCGCGCCGCGTCGAGCATCTGAGGCCC 3927

Qy        373 GAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGAC 432
Db       3926 CGCTTCCAGGAGCTGGCCGACACCCTCGTCGACGCCCTCCCGGAAACCGGCCCCGCGGAC 3867

Qy        433 GTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTT 492
Db       3866 CTGGTCACCGGCTTCGCCGCCCGCTTCCCCTTGCAGGTCCTCGCCGAAGTCATCGGCCTG 3807

Qy        493 CCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGC 552
Db       3806 CCACCGGAGTCGGCGGCCCGCTTCGACCGCGACTGGGGCAAGGTCGTCCAGCCGGTCGGA 3747
```

```
Qy        553 GTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAG 612
Db       3746 CCG-----------------ACCGACCCCGGACGCCCGCTGTACGAGGCCCGCCTGCAC 3705

Qy        613 GGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGAC 672
Db       3704 GGGCTCCAGAGCTACATAGCCGAGGTCGTCG---CCCATAAGCGCGAACACGGGGAGGAC 3648

Qy        673 GTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTG 732
Db       3647 GACCTGCTCAGCCGCCTGGTCGTGGCCCGCGACCGCAGCGAACTCTCCCAGGAGGAGCTG 3588

Qy        733 GTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCG 792
Db       3587 GACTCGATGATCTTCCAGCTCCTCGTGGCGGGCCAGGAACCGGTCACCAACCAGATCACC 3528

Qy        793 TTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCC 852
Db       3527 ACCGCCCTGATCGCCCTCTTCCGCCACCCCGCCGAACTCGCCCGGCTGCGCGACGACCCG 3468

Qy        853 GGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAATAGGA 912
Db       3467 GAGCTGATGCCCCGCGCGGTCGACGAACTCCTCCGCTACGACAGCGCCTTCGAGCTGACC 3408

Qy        913 ACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGGAG 972
Db       3407 ACCTGGCGCTTCTTCGACCAGGACAGCGACCTGCACGGCACGGACGTCCCGGCCGGGGAC 3348

Qy        973 ATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTCTCCAGGCCAGAC 1032
Db       3347 TCGGTGATCATCTCCCTGTGCGCCGCCAACCGCGACCCACGCCGCTTCCCCGACCCCGAC 3288

Qy       1033 GTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCCCCATGTC 1092
Db       3287 ACCCTCGACCTGGACCGCTCACCCAACCCCCACCTGGCCTTCGGTCACGGCATCCACTTC 3228

Qy       1093 TGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCACCATCTTCCGT 1152
Db       3227 TGCCCCGGCGCGGCCCTGGCCCGCGCCGAACTCCAGATCGCGCTGGGCACCCTCCTGTCC 3168

Qy       1153 AGGTTCCCCGAGATGAAGCTG 1173
Db       3167 CGCCTCCCCGGAGCGCACCTG 3147
```

RESULT 6
US-10-297-465B-1/c
; Sequence 1, Application US/10297465B
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Medianis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465B
; CURRENT FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465B-1

```
  Query Match             5.4%;  Score 68.6;  DB 6;  Length 2731748;
  Best Local Similarity   46.2%;  Pred. No. 1.4e-07;
  Matches  264;  Conservative    0;  Mismatches  304;  Indels    3;  Gaps    1;
```

```
Qy        602 CCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGC 661
Db     388555 CTGCCTACGAAGAGCTTGCGCAATACTTTACGAAGTTGATTGAGCTACGCCGCACCCATC 388496

Qy        662 TCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCA 721
Db     388495 CTGGAACTGACTTAATTTCTATGTTTCTGCGTGCTGAGGAAGACGGCGAGAAACTAACCC 388436

Qy        722 CGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCT 781
Db     388435 ATGATGAAATCGTCTCCAATGTGATTATGTTATTAATCGCCGGTTACGAAACCACATCCA 388376

Qy        782 ACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGA 841
Db     388375 ATATGATTGGCAATGCATTGATTGCGTTGCATCGCCATCCAGAGCAACTCGCACTCCTCA 388316

Qy        842 AGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCC 901
Db     388315 AGAGTGATCTGTCACTGATGCCACAAGCGGTATCGGAATGCTTACGTTACGACGGGTCGG 388256

Qy        902 TCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCA 961
Db     388255 TACAGTTC---ACGATGCGCGCGGCGATGGATGATATAGAGGTTGAAGGCGAGTTGGTCC 388199

Qy        962 AGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTCT 1021
Db     388198 CTCGTGGCACAGTAGTATTTTTGATGCTTGGTGCTGCCAACCGTGATCCGGCGCAATTCA 388139

Qy       1022 CCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAG 1081
Db     388138 CTCATCCCGACCAGCTGGATATTACTCGTAAGCAAGGACGCCTCCAGTCATTCGGTGCAG 388079

Qy       1082 GCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCA 1141
Db     388078 GCATTCATCATTGTCTTGGCTATCGGCTTGCGCTGATTGAACTGGAATGTGCATTGACTA 388019

Qy       1142 CCATCTTCGGTAGGTTCCCCGAGATGAAGCT 1172
Db     388018 CGCTGTTTGAACGCCTACCGCATTTAAGGCT 387988
```

RESULT 7
US-10-297-465B-1
; Sequence 1, Application US/10297465B
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Medianis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465B
; CURRENT FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465B-1

```
  Query Match             4.4%;  Score 55.8;  DB 6;  Length 2731748;
  Best Local Similarity   45.4%;  Pred. No. 0.0004;
  Matches  241;  Conservative    0;  Mismatches  287;  Indels    3;  Gaps    1;

Qy        643 GAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCC 702
```

```
Db     363970 GAGCTGCGCCGCGCGCAGCCTGGAACTGACTTGATTTCCATGTTGCTGCGTGCTGAGGAG 364029

Qy        703 GACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCT 762

Db     364030 GAGGGCCAGAAACTGACGCATGATGAAATCGTGTCCAATGTGATTTTGTTGTTATTGGGT 364089

Qy        763 GGCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCC 822

Db     364090 GGCTACGAGACCACATCCAACATGATTGGCAATGCATTGATTGCGTTGCATCGTCATCCG 364149

Qy        823 GAGGCGCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTG 882

Db     364150 AAGCAGCTTGCACGACTCAAGAGTGATTTGTCGCTGATGCCGCAAGCGATATTGGAATGC 364209

Qy        883 CTCCGCTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAG 942

Db     364210 TTACGTTACGATGGATCGGTACAGTTC---ACGATGCGCGCGGCGATGGATGATGTGAGC 364266

Qy        943 TACTGCGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTG 1002

Db     364267 ATTGAAGGTGATGTAGTACCTCGTGGCACGATTGTGTTTTGATGCTTGGTGCTGCTAAC 364326

Qy       1003 AGAGATGGGACTGTATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCG 1062

Db     364327 CGTGATCCGGCGCAATTCACTGATCCGGATCATCTGGACATCACACGGAAGCAAGGACGC 364386

Qy       1063 AGCCTCGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAG 1122

Db     364387 CTCCAGTCATTCGGCGCAGGCGTTCATCATTGCCTTGGATATCGACTTGCGTTGGTTGAG 364446

Qy       1123 GCGGAGATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTG 1173

Db     364447 TTGGAATGCGCATTGACCGTGTTGTTGGAGCGTCTGCCGCATTTGAGGTTG 364497
```

RESULT 8
US-60-501-821-38
; Sequence 38, Application US/60501821
; GENERAL INFORMATION:
; APPLICANT: ECOPIA BIOSCIENCES INC.
; APPLICANT: McAlpine, James
; TITLE OF INVENTION: POLYENE POLYKETIDES AND METHODS OF PRODUCTION
; FILE REFERENCE: 3010-1P
; CURRENT APPLICATION NUMBER: US/60/501,821
; CURRENT FILING DATE: 2003-09-11
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38
; LENGTH: 11505
; TYPE: DNA
; ORGANISM: Amycolatopsis orientalis
US-60-501-821-38

```
  Query Match             4.3%;  Score 54;  DB 7;  Length 11505;
  Best Local Similarity   45.0%;  Pred. No. 0.00021;
  Matches  244;  Conservative    0;  Mismatches  295;  Indels    3;  Gaps    1;

Qy        278 ACGGATTGTTCGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGT 337

Db       9509 ACCGGGTCTTCGGCCTGTTCGACGGCGCGTTCGGACCGGTCGCGATCGCCGACCGGCGGA 9568

Qy        338 CGTTTACGTCACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGC 397

Db       9569 CGGTCACGGAAATGCCCGTGGACTGGACGTTCGCCGAAGCGGCCGCTCTGCCGGTCGTCT 9628

Qy        398 TGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAA 457

Db       9629 TCCTCACCGCCTACTACGGGCTGGTCGACCTCGGCGGGCTCCGGCCGGGGGAGAAGGTGC 9688

Qy        458 TCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCC 517

Db       9689 TGATCCACGGAGCGACCGGCGGTGTCGGCATGGCCGCGGTCCAGCTGGCCCGCCACCTCG 9748

Qy        518 GTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATG 577

Db       9749 GCGCCGAGGTGTTCGCCACGGCGAGCCCCGGCAAGTGGGAAGTGCTGCGGGGCCTCGGTT 9808

Qy        578 AGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCC 637

Db       9809 TCGACGACGAGCACATCGCCTCCTCCCGCACGCTGGACTTCG---AGGACCGGTTCGGCC 9865

Qy        638 TCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCG 697

Db       9866 GGATGGACGTCGTCCTGGACTCGCTCGCCAAGGAGTTCGTCGACGCGTCGCTGCGGCTGC 9925

Qy        698 AGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCG 757

Db       9926 TGGGCGAGGGCGGCCGGTTCGTGGAGATGGGCAAGACCGACATCCGTGACGCGGACGAGG 9985

Qy        758 CTGCTGGCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGT 817

Db       9986 TCGCGGCCGCGCATCCCGGCGTCACCTACCGCGCGTTCGACCTGCTCGACGCCGGACGGC 10045

Qy        818 CG 819

Db      10046 CG 10047
```

RESULT 9
US-60-501-821-24
; Sequence 24, Application US/60501821
; GENERAL INFORMATION:
; APPLICANT: ECOPIA BIOSCIENCES INC.
; APPLICANT: McAlpine, James
; TITLE OF INVENTION: POLYENE POLYKETIDES AND METHODS OF PRODUCTION
; FILE REFERENCE: 3010-1P
; CURRENT APPLICATION NUMBER: US/60/501,821
; CURRENT FILING DATE: 2003-09-11
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 73599
; TYPE: DNA
; ORGANISM: Amycolatopsis orientalis
US-60-501-821-24

```
  Query Match             4.3%;  Score 54;  DB 7;  Length 73599;
  Best Local Similarity   45.0%;  Pred. No. 0.00038;
  Matches  244;  Conservative    0;  Mismatches  295;  Indels    3;  Gaps    1;

Qy        278 ACGGATTGTTCGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGT 337

Db      17912 ACCGGGTCTTCGGCCTGTTCGACGGCGCGTTCGGACCGGTCGCGATCGCCGACCGGCGGA 17971

Qy        338 CGTTTACGTCACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGC 397

Db      17972 CGGTCACGGAAATGCCCGTGGACTGGACGTTCGCCGAAGCGGCCGCTCTGCCGGTCGTCT 18031

Qy        398 TGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAA 457

Db      18032 TCCTCACCGCCTACTACGGGCTGGTCGACCTCGGCGGGCTCCGGCCGGGGGAGAAGGTGC 18091

Qy        458 TCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCC 517

Db      18092 TGATCCACGGAGCGACCGGCGGTGTCGGCATGGCCGCGGTCCAGCTGGCCCGCCACCTCG 18151

Qy        518 GTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATG 577

Db      18152 GCGCCGAGGTGTTCGCCACGGCGAGCCCCGGCAAGTGGGAAGTGCTGCGGGGCCTCGGTT 18211

Qy        578 AGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCC 637

Db      18212 TCGACGACGAGCACATCGCCTCCTCCCGCACGCTGGACTTCG---AGGACCGGTTCGGCC 18268

Qy        638 TCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCG 697
```

```
Db      18269 GGATGGACGTCGTCCTGGACTCGCTCGCCAAGGAGTTCGTCGACGCGTCGCTGCGGCTGC 18328

Qy        698 AGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCG 757
Db      18329 TGGGCGAGGGCGGCCGGTTCGTGGAGATGGGCAAGACCGACATCCGTGACGCGGACGAGG 18388

Qy        758 CTGCTGGCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGT 817
Db      18389 TCGCGGCCGCGCATCCCGGCGTCACCTACCGCGCGTTCGACCTGCTCGACGCCGGACGGC 18448

Qy        818 CG 819
Db      18449 CG 18450
```

```
RESULT 10
US-10-084-846A-77
; Sequence 77, Application US/10084846A
; GENERAL INFORMATION:
;  APPLICANT: WEITNAUER, GABRIELE
;  APPLICANT:  MUHLENWEG, AGNES
;  APPLICANT:  TREFZER, AXEL
;  APPLICANT:  BECHTHOLD, ANDREAS
;  TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
;  FILE REFERENCE: 1974-005
;  CURRENT APPLICATION NUMBER: US/10/084,846A
;  CURRENT FILING DATE:  2003-02-25
;  PRIOR APPLICATION NUMBER: PCT/EP01/09815
;  PRIOR FILING DATE: 2001-08-24
;  PRIOR APPLICATION NUMBER: DE 101 09 166.4
;  PRIOR FILING DATE: 2001-02-25
;  NUMBER OF SEQ ID NOS: 120
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 77
;   LENGTH: 972
;   TYPE: DNA
;   ORGANISM: Streptomyces viridochromogenes
;   FEATURE:
;   OTHER INFORMATION: aviABC1 dna: partial sequence of coding strand 1; nucleotide 1
;   OTHER INFORMATION: corresponds to nucleotide 34,188 of coding strand 1.
US-10-084-846A-77

  Query Match            3.9%;  Score 49.4;  DB 6;  Length 972;
  Best Local Similarity  46.5%;  Pred. No. 0.0017;
  Matches  195;  Conservative    0;  Mismatches  221;  Indels    3;  Gaps    1;

Qy        288 CGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTC 347
Db        504 CGGCCTCGACCCGCGCAACCGCAACGAGGTCTGGAAGATGGTCCGTTCGATGGTCGAGCA 563

Qy        348 ACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGC 407
Db        564 GGGCACCACCGTGCTGCTCACCACGCAGTACCTGGAGGAGGCCGACCGGCTCGCCGACCA 623

Qy        408 TCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCG 467
Db        624 GATCTCGGTGATCGACGCGGGCCGCGGGGTCGTCGAGGGCACCCCCGAGACCCTGAAGTC 683

Qy        468 CGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGG 527
Db        684 CCTGGTGGGCGGCGACCGGATCGTGGTGACCCTGGCGGCCACCGAGGAC---CTGTCGGC 740

Qy        528 CTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAA 587
Db        741 CGCCGCGGCTCAGGTCGCCCGCTTCACGGGGACCGAGCCCGAGGTGGACGCCGAGGCCCG 800

Qy        588 GACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCG 647
Db        801 CCGTATCGACGCGCCGGTACGCGACCGGGTCGGCGACCTCGTGGAGATCGTGCAGCTGCT 860

Qy        648 GCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACG 706
Db        861 GCGCGGGCAGGGCCACGGCGTGGTCGATGTGTCGCTGCGCCGCCCGACGCTGGATGACG 919
```

```
RESULT 11
US-10-084-846A-1
; Sequence 1, Application US/10084846A
; GENERAL INFORMATION:
;  APPLICANT: WEITNAUER, GABRIELE
;  APPLICANT:  MUHLENWEG, AGNES
;  APPLICANT:  TREFZER, AXEL
;  APPLICANT:  BECHTHOLD, ANDREAS
;  TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
;  FILE REFERENCE: 1974-005
;  CURRENT APPLICATION NUMBER: US/10/084,846A
;  CURRENT FILING DATE:  2003-02-25
;  PRIOR APPLICATION NUMBER: PCT/EP01/09815
;  PRIOR FILING DATE: 2001-08-24
;  PRIOR APPLICATION NUMBER: DE 101 09 166.4
;  PRIOR FILING DATE: 2001-02-25
;  NUMBER OF SEQ ID NOS: 120
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
;   LENGTH: 59816
;   TYPE: DNA
;   ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-1

  Query Match            3.9%;  Score 49.4;  DB 6;  Length 59816;
  Best Local Similarity  46.5%;  Pred. No. 0.0063;
  Matches  195;  Conservative    0;  Mismatches  221;  Indels    3;  Gaps    1;

Qy        288 CGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTC 347
Db      34691 CGGCCTCGACCCGCGCAACCGCAACGAGGTCTGGAAGATGGTCCGTTCGATGGTCGAGCA 34750

Qy        348 ACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGC 407
Db      34751 GGGCACCACCGTGCTGCTCACCACGCAGTACCTGGAGGAGGCCGACCGGCTCGCCGACCA 34810

Qy        408 TCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCG 467
Db      34811 GATCTCGGTGATCGACGCGGGCCGCGGGGTCGTCGAGGGCACCCCCGAGACCCTGAAGTC 34870

Qy        468 CGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGG 527
Db      34871 CCTGGTGGGCGGCGACCGGATCGTGGTGACCCTGGCGGCCACCGAGGAC---CTGTCGGC 34927

Qy        528 CTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAA 587
Db      34928 CGCCGCGGCTCAGGTCGCCCGCTTCACGGGGACCGAGCCCGAGGTGGACGCCGAGGCCCG 34987

Qy        588 GACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCG 647
Db      34988 CCGTATCGACGCGCCGGTACGCGACCGGGTCGGCGACCTCGTGGAGATCGTGCAGCTGCT 35047

Qy        648 GCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACG 706
Db      35048 GCGCGGGCAGGGCCACGGCGTGGTCGATGTGTCGCTGCGCCGCCCGACGCTGGATGACG 35106
```

```
RESULT 12
US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; GENERAL INFORMATION:
;  APPLICANT: WEITNAUER, GABRIELE
;  APPLICANT:  MUHLENWEG, AGNES
;  APPLICANT:  TREFZER, AXEL
;  APPLICANT:  BECHTHOLD, ANDREAS
;  TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
;  FILE REFERENCE: 1974-005
;  CURRENT APPLICATION NUMBER: US/10/084,846A
;  CURRENT FILING DATE:  2003-02-25
;  PRIOR APPLICATION NUMBER: PCT/EP01/09815
;  PRIOR FILING DATE: 2001-08-24
```

```
;  PRIOR APPLICATION NUMBER: DE 101 09 166.4
;  PRIOR FILING DATE: 2001-02-25
;  NUMBER OF SEQ ID NOS: 120
;  SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 2
;   LENGTH: 59816
;   TYPE: DNA
;   ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-2

  Query Match             3.9%;  Score 49.4;  DB 6;  Length 59816;
  Best Local Similarity   46.5%;  Pred. No. 0.0063;
  Matches  195;  Conservative    0;  Mismatches  221;  Indels    3;  Gaps    1;

Qy        288 CGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTC 347
Db      25126 CGGCCTCGACCCGCGCAACCGCAACGAGGTCTGGAAGATGGTCCGTTCGATGGTCGAGCA 25067

Qy        348 ACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGC 407
Db      25066 GGGCACCACCGTGCTGCTCACCACGCAGTACCTGGAGGAGGCCGACCGGCTCGCCGACCA 25007

Qy        408 TCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCG 467
Db      25006 GATCTCGGTGATCGACGCGGGCCGCGGGGTCGTCGAGGGCACCCCCGAGACCCTGAAGTC 24947

Qy        468 CGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGG 527
Db      24946 CCTGGTGGGCGGCGACCGGATCGTGGTGACCCTGGCGGCCACCGAGGAC---CTGTCGGC 24890

Qy        528 CTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAA 587
Db      24889 CGCCGCGGCTCAGGTCGCCCGCTTCACGGGGACCGAGCCCGAGGTGGACGCCGAGGCCCG 24830

Qy        588 GACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCG 647
Db      24829 CCGTATCGACGCGCCGGTACGCGACCGGGTCGGCGACCTCGTGGAGATCGTGCAGCTGCT 24770

Qy        648 GCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACG 706
Db      24769 GCGCGGGCAGGGCCACGGCGTGGTCGATGTGTCGCTGCGCCGCCCGACGCTGGATGACG 24711
```

```
RESULT 13
PCT-US03-11231-190
; Sequence 190, Application PC/TUS0311231
; GENERAL INFORMATION:
;  APPLICANT: Corixa Corporation
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Hosken, Nancy A.
;  APPLICANT:  Parsons, Joseph M.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION:  TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
;  FILE REFERENCE: 210121.53801PC
;  CURRENT APPLICATION NUMBER: PCT/US03/11231
;  CURRENT FILING DATE:  2003-04-09
;  NUMBER OF SEQ ID NOS: 267
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 190
;   LENGTH: 9369
;   TYPE: DNA
;   ORGANISM: HSV2
PCT-US03-11231-190

  Query Match             3.9%;  Score 48.8;  DB 1;  Length 9369;
  Best Local Similarity   46.1%;  Pred. No. 0.005;
  Matches  201;  Conservative    0;  Mismatches  232;  Indels    3;  Gaps    1;

Qy        313 CGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGCGCGCC 372
Db       4489 CGCGATCGCCTGGACGCCATCCGGGCCGACGTGCACCGGGCGCTCGGCGGCGTGCCCCTG 4548

Qy        373 GAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGAC 432
Db       4549 GACCTGGCGGCCGCCGCGGAGCAGATATCCGCGGCCCGGAATGACCCGGAGGCCACCGCG 4608

Qy        433 GTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTT 492
Db       4609 GAGCTGGTCCGGACCGGGGTCACCCTGCCGTGTCCCAGCGAGGACGCGCTGGTCGCGTGC 4668

Qy        493 CCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGC 552
Db       4669 GCGGCGGCGTTGGAGCGCGTGGACCAGAGCCCCGTGAAGAACACGGCGTACGCCGAGTAT 4728

Qy        553 GTGGGTTTGGTGCCC---CAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACC 609
Db       4729 GTCGCCTTCGTGACCCGGCAGGACACGGCCGAGACCAAGGACGCAGTGGTGCGCGCCAAA 4788

Qy        610 GAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAAT 669
Db       4789 CAGCAGCGCGCGGAGGCCACCGAGCGGGTCATGGCGGGCCTGCGGGAGGCGCTGGCCGCG 4848

Qy        670 GACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAG 729
Db       4849 CGCGAGCGCCGGGCCCAGATCGAGGCCGAGGGCCTGGCCAACCTCAAGACCATGCTGAAG 4908

Qy        730 CTGGTCGCGCTCGTGG 745
Db       4909 GTGGTCGCCGTCCCGG 4924
```

```
RESULT 14
PCT-US03-11231-247
; Sequence 247, Application PC/TUS0311231
; GENERAL INFORMATION:
;  APPLICANT: Corixa Corporation
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Hosken, Nancy A.
;  APPLICANT:  Parsons, Joseph M.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION:  TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
;  FILE REFERENCE: 210121.53801PC
;  CURRENT APPLICATION NUMBER: PCT/US03/11231
;  CURRENT FILING DATE:  2003-04-09
;  NUMBER OF SEQ ID NOS: 267
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 247
;   LENGTH: 9369
;   TYPE: DNA
;   ORGANISM: Herpes simplex virus
PCT-US03-11231-247

  Query Match             3.9%;  Score 48.8;  DB 1;  Length 9369;
  Best Local Similarity   46.1%;  Pred. No. 0.005;
  Matches  201;  Conservative    0;  Mismatches  232;  Indels    3;  Gaps    1;

Qy        313 CGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGCGCGCC 372
Db       4489 CGCGATCGCCTGGACGCCATCCGGGCCGACGTGCACCGGGCGCTCGGCGGCGTGCCCCTG 4548

Qy        373 GAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGAC 432
Db       4549 GACCTGGCGGCCGCCGCGGAGCAGATATCCGCGGCCCGGAATGACCCGGAGGCCACCGCG 4608

Qy        433 GTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTT 492
Db       4609 GAGCTGGTCCGGACCGGGGTCACCCTGCCGTGTCCCAGCGAGGACGCGCTGGTCGCGTGC 4668

Qy        493 CCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGC 552
Db       4669 GCGGCGGCGTTGGAGCGCGTGGACCAGAGCCCCGTGAAGAACACGGCGTACGCCGAGTAT 4728

Qy        553 GTGGGTTTGGTGCCC---CAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACC 609
Db       4729 GTCGCCTTCGTGACCCGGCAGGACACGGCCGAGACCAAGGACGCAGTGGTGCGCGCCAAA 4788
```

```
Qy        610 GAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAAT 669

Db       4789 CAGCAGCGCGCGGAGGCCACCGAGCGGGTCATGGCGGGCCTGCGGGAGGCGCTGGCCGCG 4848

Qy        670 GACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAG 729

Db       4849 CGCGAGCGCCGGGCCCAGATCGAGGCCGAGGGCCTGGCCAACCTCAAGACCATGCTGAAG 4908

Qy        730 CTGGTCGCGCTCGTGG 745

Db       4909 GTGGTCGCCGTCCCGG 4924


RESULT 15
US-10-425-114A-14511
; Sequence 14511, Application US/10425114A
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114A
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 14511
;   LENGTH: 1036
;   TYPE: DNA
;   ORGANISM: Chlorella vulgaris
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB191-016-B10_FLI
US-10-425-114A-14511

  Query Match             3.8%;  Score 48;  DB 6;  Length 1036;
  Best Local Similarity   48.9%;  Pred. No. 0.0041;
  Matches  129;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        448 GCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGAC 507

Db        116 GGGGAGGAGTTTGGCATCATTGCCACCTTCGCGCCCAAGCCCATGAAGGGCGACTGGAAC 175

Qy        508 GAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCC 567

Db        176 GGCACCGGCGCGCACACCAACTACTCCACGGAGAGCATGCGCAAGGACGGCGGCATGCGT 235

Qy        568 CAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTC 627

Db        236 GCGATTGAGGACGCGATCGAGAAGCTGTCCAAGTGCCACCCGGAGCACATTGCGCAGTAC 295

Qy        628 CATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTG 687

Db        296 GGGTACGGCAACGAGGAGCGGCTGACCGGCAAGCACGAGACATGCGACATCAACACCTTC 355

Qy        688 CTTCAGGCCGAGGCCGACGGCAGC 711

Db        356 CGCTACGGCGTGGCCGACCGCGGC 379


Search completed: October  4, 2003, 12:01:00
Job time : 233.454 secs
```

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 03:31:13 ; Search time 289.856 Seconds
(without alignments)
11088.586 Million cell updates/sec

Title: US-10-014-717-1_COPY_62369_63628
Perfect score: 1260
Sequence: 1 atgacacaggagcaagcgaa..........agccctccaaagctggatag 1260

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1260	100.0	68750	13	US-10-014-717-1	Sequence 1, Appli
	2	151.4	12.0	1227	12	US-10-214-446-3	Sequence 3, Appli
	3	149.4	11.9	1233	14	US-10-205-032-7	Sequence 7, Appli
c	4	149.4	11.9	60196	14	US-10-205-032-1	Sequence 1, Appli
	5	148.8	11.8	1164	14	US-10-156-761-6226	Sequence 6226, Ap
	6	148.8	11.8	9025608	14	US-10-156-761-1	Sequence 1, Appli
	7	146.4	11.6	1197	14	US-10-156-761-409	Sequence 409, App
	8	146.4	11.6	100000	14	US-10-156-761-15103	Sequence 15103, A
c	9	146.4	11.6	9025608	14	US-10-156-761-1	Sequence 1, Appli
	10	142.6	11.3	1215	12	US-10-214-446-49	Sequence 49, Appl
	11	135	10.7	1248	12	US-10-214-446-55	Sequence 55, Appl
	12	134.8	10.7	1248	12	US-10-214-446-45	Sequence 45, Appl
	13	132	10.5	1197	14	US-10-156-761-2364	Sequence 2364, Ap
	14	131.2	10.4	1179	12	US-10-214-446-31	Sequence 31, Appl
	15	129.4	10.3	1278	12	US-10-214-446-19	Sequence 19, Appl
	16	127	10.1	1179	14	US-10-156-761-1975	Sequence 1975, Ap
	17	123.4	9.8	1248	14	US-10-156-761-7447	Sequence 7447, Ap
	18	123	9.8	1251	10	US-09-861-289-38	Sequence 38, Appl
	19	123	9.8	1251	10	US-09-860-846-38	Sequence 38, Appl
	20	123	9.8	1251	11	US-09-988-384B-38	Sequence 38, Appl
	21	123	9.8	1251	11	US-09-836-821-38	Sequence 38, Appl
	22	123	9.8	5970	11	US-09-793-708-21	Sequence 21, Appl
	23	123	9.8	5970	12	US-10-201-365-11	Sequence 11, Appl
	24	123	9.8	5970	12	US-10-160-539-21	Sequence 21, Appl
	25	120.4	9.6	1191	14	US-10-156-761-7165	Sequence 7165, Ap
	26	120	9.5	1215	12	US-10-214-446-39	Sequence 39, Appl
	27	119.8	9.5	1209	14	US-10-205-032-9	Sequence 9, Appli
	28	119.8	9.5	60196	14	US-10-205-032-1	Sequence 1, Appli
	29	119.6	9.5	1215	14	US-10-156-761-4523	Sequence 4523, Ap
	30	118	9.4	1212	14	US-10-156-761-2881	Sequence 2881, Ap
c	31	118	9.4	125746	14	US-10-156-761-15102	Sequence 15102, A
	32	116	9.2	1221	12	US-10-214-446-1	Sequence 1, Appli
	33	115	9.1	987	12	US-10-132-134-24	Sequence 24, Appl
	34	115	9.1	1248	14	US-10-156-761-408	Sequence 408, App
	35	115	9.1	1700	12	US-10-132-134-22	Sequence 22, Appl
	36	114.6	9.1	1428	14	US-10-145-415-15	Sequence 15, Appl
	37	114.4	9.1	1233	12	US-09-953-348-24	Sequence 24, Appl
	38	114.4	9.1	1233	14	US-10-267-255-24	Sequence 24, Appl
	39	114.4	9.1	12249	12	US-09-953-348-74	Sequence 74, Appl
	40	114.4	9.1	12249	14	US-10-267-255-74	Sequence 74, Appl
	41	114.4	9.1	18331	12	US-09-953-348-96	Sequence 96, Appl
	42	114.4	9.1	18331	14	US-10-267-255-96	Sequence 96, Appl
	43	111.8	8.9	1293	14	US-10-145-415-7	Sequence 7, Appli
	44	111.8	8.9	1293	14	US-10-145-415-19	Sequence 19, Appl
	45	111.4	8.8	1428	14	US-10-145-415-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-10-014-717-1
; Sequence 1, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-014-717-1

```
  Query Match             100.0%;  Score 1260;  DB 13;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy       181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
Db     62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy       241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
Db     62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy       301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
Db     62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy       361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
Db     62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy       421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
Db     62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy       481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
Db     62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy       541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
Db     62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy       601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
Db     62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy       661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
Db     63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy       721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
Db     63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy       781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
Db     63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy       841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
Db     63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy       901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
Db     63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy       961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
Db     63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy      1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
Db     63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy      1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
Db     63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy      1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
Db     63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy      1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
Db     63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628
```

```
RESULT 2
US-10-214-446-3
; Sequence 3, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
;  APPLICANT: Weiner, David
;  APPLICANT:  Burk, Mark J.
;  APPLICANT:  Hitchman, Tim
;  APPLICANT:  Pujol, Catherine
;  APPLICANT:  Richardson, Toby
;  APPLICANT:  Short, Jay M.
;  TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
;  TITLE OF INVENTION:  THEM AND METHODS OF MAKING AND USING THEM
;  FILE REFERENCE: 09010-500001
;  CURRENT APPLICATION NUMBER: US/10/214,446
;  CURRENT FILING DATE:  2002-08-05
;  PRIOR APPLICATION NUMBER: US 60/309,497
;  PRIOR FILING DATE: 2001-08-03
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1227
;   TYPE: DNA
;   ORGANISM: Bacterial
US-10-214-446-3

  Query Match             12.0%;  Score 151.4;  DB 12;  Length 1227;
  Best Local Similarity   49.9%;  Pred. No. 5.7e-34;
  Matches  513;  Conservative    0;  Mismatches  486;  Indels   30;  Gaps    4;

Qy       147 CTGGGTCCTCACCCGATACCACGACGTGTCGGCGGTGTTCCGCGACGAACGCTTCGCGGT 206
Db       138 CTGGATGGTCACCGGATACGACGAGGCGGTGGCGGCGCTCACCGACCCCCGGCTCAGCAG 197

Qy       207 CAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTACTCGTCGGCCATTCCCGAGCTCAGCGA 266
Db       198 CAGCCCCGTCGGCGTCAACGGACTCGAGGAGGA---GATGGCCCACCAGGAGCGCACCAA 254

Qy       267 TATGAAGAAGTACGGATTGTTCGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCT 326
Db       255 CGTCCTGATGGCCAGCATGCTCGTCGCCAACGGCGAGGACCACACCCGGCTGCGCAACCT 314

Qy       327 CGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCAC 386
Db       315 CGTCTCGAAGGCCTTCACCGCCCGCCGGGTGGAGCAGCTCGCGCCGCGCGTCCAGGCGCA 374

Qy       387 CGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTA 446
Db       375 CACCGACGCCTTCCTCGACGCGGTCGCGGCGCGCGGATCCGCCGACCTGGTCTCGGAGTT 434

Qy       447 CGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGA 506
Db       435 CGCCCTGCCGCTCCCCATGGCCGTACTCAGCGACCTCATCGGCATCCCGGCCGAGGGGCA 494

Qy       507 CGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCC 566
Db       495 GCCCGACTTCGCCCGC-----------------CTCGCGGTCGGCCTCATCATGCCGCC 536

Qy       567 CCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCT 626
Db       537 GAACACCCCCGAGCGGCTCGCCAAGGGAGCCCGCGCCCGCGCCGAACTCACCGAGTTCTT 596

Qy       627 CCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCT 686
Db       597 CGAGCCGTTGATCGCCCAGCGCAAGAAGGAGCCG------AAGGACGACCTGCTGAGCGC 650

Qy       687 GCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGG 746
Db       651 GCTCTGCGCGGCGCAGGCCGAGGAGCGGATCAGCGACCGCGAGCTGACGGCGATGACGAT 710

Qy       747 TGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAA 806
```

```
Db        711 CCTGCTCACGCTCGCCGGGCACGAGACGACGGCCAGCCTGATCGCCAACGGCGTGCACGC 770

Qy        807 CCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAA 866

Db        771 CCTGCTGCGGCACCCGGAGCAGTTCGCCACCCTGCGCGACGACCCCTCGCTGCTGCCGGG 830

Qy        867 CGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGC 926

Db        831 CGCGATCGAGGAACTCCTGCGCTACGAGGGCCCGGTGAGCCGGGGCGTCGCCCGCTTCAC 890

Qy        927 CAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCT 986

Db        891 CACCGACCCGTACGAGATCGGCGGGGTCACCGTACCGGCCGGCGAGATGATCATCATCGG 950

Qy        987 GATCCCGAGCGCCCTGAGAGATGGGACTGTATTCTCCAGGCCAGACGTGTTTGA---TGT 1043

Db        951 GCTCGCCGCGGCCAATCGCGACCCGGCCCGCTACGACCGTCCCGACATCCTCGACGTTGC 1010

Qy       1044 GCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGT 1103

Db       1011 ACGCCGTGAGGTGCCGCAACAGCTCGCTTTCGGCCATGGCGTGCACTTCTGCCTGGGTGC 1070

Qy       1104 GTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGA 1163

Db       1071 GCCGCTGGCCCGCGCGGAGGCCCGGATCGCCATCGGCACCCTGCTGCGCCGCTTCCCCGA 1130

Qy       1164 GATGAAGCT 1172

Db       1131 TCTGCGGCT 1139
```

RESULT 3
US-10-205-032-7
; Sequence 7, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-2US
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-7

```
  Query Match             11.9%;  Score 149.4;  DB 14;  Length 1233;
  Best Local Similarity   49.7%;  Pred. No. 2.2e-33;
  Matches  437;  Conservative    0;  Mismatches  436;  Indels    6;  Gaps    2;

Qy        288 CGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTC 347

Db        261 CGGCATCGACCCGCCGCGGCACGCGCCGCTGCGCAAGCTGGTGAGCCAGGCGTTCACCCC 320

Qy        348 ACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGC 407

Db        321 CCGCCGGATCGCCCAGATGGAGCTGCGGATCGGGCAGATCACCGCCGACGTGCTCGACCA 380

Qy        408 TCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCG 467

Db        381 GGTACGCGACCAGGACCGGATCGACATCGCCAGCGACCTCGCGTACCCGCTGCCGGTGAC 440

Qy        468 CGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGG 527

Db        441 GGTCATCGCCGAGCTGCTCGGCATTCCCACCAAGGATCACGAGAAGTTCCGCGAGTGGGT 500

Qy        528 CTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAA 587

Db        501 GGACATCATCCTCAGCAACGAAGGGCTGGAGTATCCCAACCTCCCGGACGACTTCACCGA 560

Qy        588 GACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCG 647

Db        561 GACGGTGGGCCCCGCCATCGAGGAGTGGTCCGAATTCCTGTACGCCCAGATCGCCCACAA 620

Qy        648 GCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGG 707

Db        621 GCGCGCCGAACCG---AAGGACGACCTGATCAGCGGCCTCTGTGCGGCGGAGGTCGACGG 677

Qy        708 CAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCAC 767

Db        678 GCGCAAGCTGACCGACGAGGAAGTCGTCAACATCGTCGCGCTGCTGCTCACCGCCGGGCA 737

Qy        768 CGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGC 827

Db        738 CATCTCCAGCGCCACGCTGCTCAGCAACCTGTTCCTGGTGCTGGAGGAGCACCCGCAGGC 797

Qy        828 GCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCG 887

Db        798 ACAGGCCGCGGTCCGCGCCGACCGCAGCCTCGTGCCGGGCGTGATCGAGGAGACGCTGCG 857

Qy        888 CTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTG 947

Db        858 CT---ACCGGTCCCCGTTCAACTGCATCTTCCGGATCCTGAACGAGGACACCGACATCCT 914

Qy        948 CGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGA 1007

Db        915 CGGCCACCCCATGCGCAAGGGCCAGATGGTGATCGCCTGGATCGCCTCCGCGAACCGCGA 974

Qy       1008 TGGGACTGTATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCT 1067

Db        975 CACCGAGGTGTTCACGGACCCGGACACCTTCGACATCCGACGCGAGTCGAACAAGCACCT 1034

Qy       1068 CGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGA 1127

Db       1035 GGCGTTCGGCCACGGCATCCACCACTGCCTGGGCGCGTTCCTGGCCAGGCTGGAGGCGAA 1094

Qy       1128 GATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGAT 1166

Db       1095 GGTCTTCCTCAACCAGACGCTCGACCAGTTCACCGAGTT 1133
```

RESULT 4
US-10-205-032-1/c
; Sequence 1, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-2US
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 60196
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-1

```
  Query Match             11.9%;  Score 149.4;  DB 14;  Length 60196;
  Best Local Similarity   49.7%;  Pred. No. 4.3e-33;
  Matches  437;  Conservative    0;  Mismatches  436;  Indels    6;  Gaps    2;

Qy        288 CGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTC 347

Db       3601 CGGCATCGACCCGCCGCGGCACGCGCCGCTGCGCAAGCTGGTGAGCCAGGCGTTCACCCC 3542

Qy        348 ACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGC 407
```

```
Db       3541 CCGCCGGATCGCCCAGATGGAGCTGCGGATCGGGCAGATCACCGCCGACGTGCTCGACCA 3482

Qy        408 TCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCG 467

Db       3481 GGTACGCGACCAGGACCGGATCGACATCGCCAGCGACCTCGCGTACCCGCTGCCGGTGAC 3422

Qy        468 CGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGG 527

Db       3421 GGTCATCGCCGAGCTGCTCGGCATTCCCACCAAGGATCACGAGAAGTTCCGCGAGTGGGT 3362

Qy        528 CTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAA 587

Db       3361 GGACATCATCCTCAGCAACGAAGGGCTGGAGTATCCCAACCTCCCGGACGACTTCACCGA 3302

Qy        588 GACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCG 647

Db       3301 GACGGTGGGCCCCGCCATCGAGGAGTGGTCCGAATTCCTGTACGCCCAGATCGCCCACAA 3242

Qy        648 GCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGG 707

Db       3241 GCGCGCCGAACCG---AAGGACGACCTGATCAGCGGCCTCTGTGCGGCGGAGGTCGACGG 3185

Qy        708 CAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCAC 767

Db       3184 GCGCAAGCTGACCGACGAGGAAGTCGTCAACATCGTCGCGCTGCTGCTCACCGCCGGGCA 3125

Qy        768 CGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGC 827

Db       3124 CATCTCCAGCGCCACGCTGCTCAGCAACCTGTTCCTGGTGCTGGAGGAGCACCCGCAGGC 3065

Qy        828 GCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCG 887

Db       3064 ACAGGCCGCGGTCCGCGCCGACCGCAGCCTCGTGCCGGGCGTGATCGAGGAGACGCTGCG 3005

Qy        888 CTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTG 947

Db       3004 CT---ACCGGTCCCCGTTCAACTGCATCTTCCGGATCCTGAACGAGGACACCGACATCCT 2948

Qy        948 CGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGA 1007

Db       2947 CGGCCACCCCATGCGCAAGGGCCAGATGGTGATCGCCTGGATCGCCTCCGCGAACCGCGA 2888

Qy       1008 TGGGACTGTATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCT 1067

Db       2887 CACCGAGGTGTTCACGGACCCGGACACCTTCGACATCCGACGCGAGTCGAACAAGCACCT 2828

Qy       1068 CGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGA 1127

Db       2827 GGCGTTCGGCCACGGCATCCACCACTGCCTGGGCGCGTTCCTGGCCAGGCTGGAGGCGAA 2768

Qy       1128 GATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGAT 1166

Db       2767 GGTCTTCCTCAACCAGACGCTCGACCAGTTCACCGAGTT 2729
```

RESULT 5
US-10-156-761-6226
; Sequence 6226, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6226
;   LENGTH: 1164
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1164)
US-10-156-761-6226

```
  Query Match             11.8%;  Score 148.8;  DB 14;  Length 1164;
  Best Local Similarity   50.0%;  Pred. No. 3.2e-33;
  Matches  449;  Conservative    0;  Mismatches  422;  Indels   27;  Gaps    2;

Qy        284 TGTTCGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTA 343

Db        233 TGCTCAACATGGACGCGCCGGACCACACCCGCATCCGGCGCCTGGTCGGCCGGGCCTTCA 292

Qy        344 CGTCACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCG 403

Db        293 CCCTGCGTCGCGTCGAACAACTCCGCGAGCCCGTACGGGAGACCGCCCACCGGCTTCTCG 352

Qy        404 ATGCTCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGA 463

Db        353 ACGCACTCGGGACGCACGGCAGCACCGACCTCATCGCCTCGTACGCCGCGCCCCTGCCGA 412

Qy        464 TGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCT 523

Db        413 TCACCGTCATCTGCGACCTGCTCGGCGTGCCGGACGAACACCGGCGGGACTTCCGGGCCT 472

Qy        524 TCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGA 583

Db        473 GGACCGACCCGC----------------------TCGTCACCCCGGACCCGGCCCCGCC 508

Qy        584 CCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATG 643

Db        509 CGGACGTCGCCAGGGAATCCGTCGTCTCGTTGCTCGGCTTCTTCACCGGGCTCCTCGCCG 568

Qy        644 AGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCG 703

Db        569 ACAAGCGCAAGAACCC---CGCGGACGACCTGCTCTCCGACCTCATCGCCGTGCAGGAAG 625

Qy        704 ACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTG 763

Db        626 AGGGGGACCGGCTCACCGAGGACGAGCTGATGTCCCTCGCCTTCCTCATCCTCTTCGCCG 685

Qy        764 GCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCG 823

Db        686 GCTACGAGAACACCGTGCACCTCATCGGGAACGCCGTACTGGCGCTGCTGCGCCATCCGG 745

Qy        824 AGGCGCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGC 883

Db        746 AGCAGCTCGCCGCGCTCCGCGAGGACCCGGCACGGCTGCCGGACGCCGTCGGGGAGTTCG 805

Qy        884 TCCGCTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGT 943

Db        806 CCCGCTACGAAGGGCCCGCCCTGCTCGCCATCCGCCGCTTCCCGGTGCGGGACGTGACGA 865

Qy        944 ACTGCGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGA 1003

Db        866 TCGGCGGGGTCACCGTGCCCGCGGGGGAGACCGTGCTGCTGTCGCTGTCCGCCGCCAACC 925

Qy       1004 GAGATGGGACTGTATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGA 1063

Db        926 GTGACCCGAGCCGGTTCCCCGACCCCGACCGCCTCGATCTCGGCCGCGACGCCGCAGGCC 985

Qy       1064 GCCTCGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGG 1123

Db        986 ACCTCGCGCTCGGCCACGGCGTCCACTACTGCCTGGGCGCGCCGCTGGCCCGGCTGGAGA 1045
```

```
Qy       1124 CGGAGATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAAC 1181
Db       1046 CGGAGGTCGCCCTCGCCGCGCTCCTCGAACGGTTCCCGGACCTCGCCCTCGCGGAAAC 1103

RESULT 6
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

  Query Match             11.8%;  Score 148.8;  DB 14;  Length 9025608;
  Best Local Similarity   50.0%;  Pred. No. 1.5e-32;
  Matches  449;  Conservative    0;  Mismatches  422;  Indels   27;  Gaps    2;

Qy        284 TGTTCGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTA 343
Db    7508344 TGCTCAACATGGACGCGCCGGACCACACCCGCATCCGGCGCCTGGTCGGCCGGGCCTTCA 7508403

Qy        344 CGTCACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCG 403
Db    7508404 CCCTGCGTCGCGTCGAACAACTCCGCGAGCCCGTACGGGAGACCGCCCACCGGCTTCTCG 7508463

Qy        404 ATGCTCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGA 463
Db    7508464 ACGCACTCGGGACGCACGGCAGCACCGACCTCATCGCCTCGTACGCCGCGCCCCTGCCGA 7508523

Qy        464 TGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCT 523
Db    7508524 TCACCGTCATCTGCGACCTGCTCGGCGTGCCGGACGAACACCGGCGGGACTTCCGGGCCT 7508583

Qy        524 TCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGA 583
Db    7508584 GGACCGACCCGC-----------------------TCGTCACCCCGGACCCGGCCCGCC 7508619

Qy        584 CCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATG 643
Db    7508620 CGGACGTCGCCAGGGAATCCGTCGTCTCGTTGCTCGGCTTCTTCACCGGGCTCCTCGCCG 7508679

Qy        644 AGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCG 703
Db    7508680 ACAAGCGCAAGAACCC---CGCGGACGACCTGCTCTCCGACCTCATCGCCGTGCAGGAAG 7508736

Qy        704 ACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTG 763
Db    7508737 AGGGGGACCGGCTCACCGAGGACGAGCTGATGTCCCTCGCCTTCCTCATCCTCTTCGCCG 7508796

Qy        764 GCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCG 823
Db    7508797 GCTACGAGAACACCGTGCACCTCATCGGGAACGCCGTACTGGCGCTGCTGCGCCATCCGG 7508856

Qy        824 AGGCGCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGC 883
Db    7508857 AGCAGCTCGCCGCGCTCCGCGAGGACCCGGCACGGCTGCCGGACGCCGTCGGGGAGTTCG 7508916

Qy        884 TCCGCTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGT 943
Db    7508917 CCCGCTACGAAGGGCCCGCCCTGCTCGCCATCCGCCGCTTCCCGGTGCGGACGTGACGA 7508976

Qy        944 ACTGCGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGA 1003
Db    7508977 TCGGCGGGGTCACCGTGCCCGCGGGGGAGACCGTGCTGCTGTCGCTGTCCGCCGCCAACC 7509036

Qy       1004 GAGATGGGACTGTATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGA 1063
Db    7509037 GTGACCCGAGCCGGTTCCCCGACCCCGACCGCCTCGATCTCGGCCGCGACGCCGCAGGCC 7509096

Qy       1064 GCCTCGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGG 1123
Db    7509097 ACCTCGCGCTCGGCCACGGCGTCCACTACTGCCTGGGCGCGCCGCTGGCCCGGCTGGAGA 7509156

Qy       1124 CGGAGATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAAC 1181
Db    7509157 CGGAGGTCGCCCTCGCCGCGCTCCTCGAACGGTTCCCGGACCTCGCCCTCGCGGAAAC 7509214

RESULT 7
US-10-156-761-409
; Sequence 409, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 409
;   LENGTH: 1197
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1197)
US-10-156-761-409

  Query Match             11.6%;  Score 146.4;  DB 14;  Length 1197;
  Best Local Similarity   50.8%;  Pred. No. 1.6e-32;
  Matches  453;  Conservative    0;  Mismatches  406;  Indels   33;  Gaps    3;

Qy        279 CGGATTGTTCGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTC 338
Db        255 CAGCCTGTTGTGGCAGGACGAGCCGGAGCACACCTCGGACCGCAAGCTGCTCGCCAAGGA 314

Qy        339 GTTTACGTCACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCT 398
Db        315 GTTCACCGTACGGCGCATGCAGGCGCTGCGGCCGAACATCCAGCGCATCGTCGACGAGCA 374

Qy        399 GCTCGATGCTCGCTCCGGACAAGAGG---AGTTCGACGTTGTGCGGGATTACGCGGAGGG 455
Db        375 CCTCGATGCCATCGAGGCGCGGGGCGGCCCGGTCGACCTGGTGAAGACCTTCGCCAACGC 434
```

```
Qy        456 AATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTT 515
Db        435 GGTGCCGTCCATGGTGATTTCCGACCTGTTCGGTGTCCCGGTCGAGCGCCGGGCGGAGTT 494

Qy        516 CCGTCGCTTCGGCTCGGCGACTGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGA 575
Db        495 CCAGGACATCGCCGAGGCGATGATGCG--------------------------GGTCGA 527

Qy        576 TGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGT 635
Db        528 CCAGGACGCCGCCGCCACCGAGGCCGCCGGCATGCGCCTGGGCGGGCTGCTCTACCAGCT 587

Qy        636 CCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGC 695
Db        588 TGTCCAGGAGCGCCGGGCCAACCC---CGGTGACGACCTGATCTCGGCGCTGATCACCAC 644

Qy        696 CGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTAT 755
Db        645 CGAGGACCCCGATGGTGTGGTCGACGACATGTTTCTCATGAACGCGGCCGGCACCCTGCT 704

Qy        756 CGCTGCTGGCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCG 815
Db        705 GATAGCGGCCCACGACACCACCGCCTGCATGATCGGCCTCGGTACGGCGCTGCTCCTCGA 764

Qy        816 GTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGA 875
Db        765 CAGCCCCGACCAGCTGGCCCTGCTGCGCGAGGACCCGTCGCTGGTCGGCAACGCCGTCGA 824

Qy        876 TGAGGTGCTCCGCTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGA 935
Db        825 GGAGCTGCTGCGCTACCTCACCATCGGCCAGTTCGGCGGCGAGCGCGTCGCCACCCGGGA 884

Qy        936 CCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAG 995
Db        885 CGTGGAGCTCGGTGGGGTGCGCATCGCCAAGGGCGAGCAGGTCGTCGCGCACGTCCTCGC 944

Qy        996 CGCCCTGAGAGATGGACTGTATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGACAC 1055
Db        945 CGCCGACTTCGACCCCGCCTTCGTGGAGGAGCCGGAACGTTTCGACATCACTCGACGCCC 1004

Qy       1056 GGGCGCGAGCCTCGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCG 1115
Db       1005 CGCCCCGCATCTGGCCTTCGGCTTCGGCGCACACCAGTGCATCGGGCAGCAACTCGCCCG 1064

Qy       1116 CCTCGAGGCGGAGATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGATG 1167
Db       1065 GATCGAGCTGCAGATCGTCTTCGAGACCCTGTTCCGCCGCCTCCCGGGTCTG 1116
```

RESULT 8
US-10-156-761-15103
; Sequence 15103, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15103
; LENGTH: 100000
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15103

Query Match 11.6%; Score 146.4; DB 14; Length 100000;
Best Local Similarity 50.8%; Pred. No. 3.5e-32;
Matches 453; Conservative 0; Mismatches 406; Indels 33; Gaps 3;

```
Qy        279 CGGATTGTTCGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTC 338
Db      80684 CAGCCTGTTGTGGCAGGACGAGCCGGAGCACACCTCGGACCGCAAGCTGCTCGCCAAGGA 80743

Qy        339 GTTTACGTCACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCT 398
Db      80744 GTTCACCGTACGGCGCATGCAGGCGCTGCGGCCGAACATCCAGCGCATCGTCGACGAGCA 80803

Qy        399 GCTCGATGCTCGCTCCGGACAAGAGG---AGTTCGACGTTGTGCGGGATTACGCGGAGGG 455
Db      80804 CCTCGATGCCATCGAGGCGCGGGGCGGCCCGGTCGACCTGGTGAAGAC
```

```
OM nucleic - nucleic search, using sw model

Run on:         October  3, 2003, 11:46:26 ; Search time 66.7507 Seconds
                                             (without alignments)
                                             8331.640 Million cell updates/sec

Title:          US-10-014-717-1_COPY_62369_63628
Perfect score:  1260
Sequence:       1 atgacacaggagcaagcgaa..........agccctccaaagctggatag 1260

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters:      1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1260	100.0	68750	3	US-09-335-409-1	Sequence 1, Appli
	2	1260	100.0	68750	4	US-09-568-102-1	Sequence 1, Appli
	3	1260	100.0	68750	4	US-09-567-969-1	Sequence 1, Appli
	4	1260	100.0	68750	4	US-09-568-480-1	Sequence 1, Appli
	5	1260	100.0	68750	4	US-09-568-486-1	Sequence 1, Appli
	6	1260	100.0	68750	4	US-09-568-472-1	Sequence 1, Appli
	7	1260	100.0	68750	4	US-09-567-899-1	Sequence 1, Appli
	8	1246.2	98.9	71989	4	US-09-443-501A-2	Sequence 2, Appli
	9	145.4	11.5	1314	4	US-09-252-991A-1265	Sequence 1265, Ap
c	10	145.4	11.5	1482	4	US-09-252-991A-1338	Sequence 1338, Ap
	11	145.4	11.5	1671	4	US-09-252-991A-1222	Sequence 1222, Ap
	12	123	9.8	1251	3	US-09-105-537-38	Sequence 38, Appl
	13	123	9.8	5970	3	US-09-320-878-21	Sequence 21, Appl
	14	123	9.8	5970	4	US-09-141-908-11	Sequence 11, Appl
	15	123	9.8	5970	4	US-09-657-440-21	Sequence 21, Appl
	16	118.2	9.4	1407	4	US-09-252-991A-15866	Sequence 15866, A
	17	118.2	9.4	1731	4	US-09-252-991A-15928	Sequence 15928, A
c	18	118.2	9.4	2502	4	US-09-252-991A-15767	Sequence 15767, A
	19	115.6	9.2	1212	6	5212296-17	Patent No. 5212296
	20	115	9.1	1221	6	5212296-16	Patent No. 5212296
	21	115	9.1	1879	6	5212296-5	Patent No. 5212296
	22	114.4	9.1	1227	3	US-09-385-028-23	Sequence 23, Appl
	23	114.4	9.1	1227	4	US-09-726-614-23	Sequence 23, Appl
	24	114.4	9.1	1233	4	US-09-266-965-24	Sequence 24, Appl
	25	114.4	9.1	12249	4	US-09-266-965-74	Sequence 74, Appl
	26	114.4	9.1	15079	3	US-09-385-028-1	Sequence 1, Appli
	27	114.4	9.1	15079	4	US-09-726-614-1	Sequence 1, Appli
	28	114.4	9.1	18331	4	US-09-266-965-96	Sequence 96, Appl
	29	111.4	8.8	1735	1	US-08-102-863-10	Sequence 10, Appl
	30	111.4	8.8	1735	5	PCT-US92-10885-10	Sequence 10, Appl
	31	109.6	8.7	1224	4	US-09-266-965-22	Sequence 22, Appl
	32	97.4	7.7	50937	3	US-09-428-517-1	Sequence 1, Appli
	33	94.8	7.5	1998	6	5212296-8	Patent No. 5212296
c	34	94.6	7.5	47981	4	US-09-679-279-1	Sequence 1, Appli
c	35	86.2	6.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	36	86.2	6.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	37	82.8	6.6	77536	4	US-09-410-551B-1	Sequence 1, Appli
	38	80.2	6.4	6085	3	US-09-029-603-4	Sequence 4, Appli
	39	77.6	6.2	1188	4	US-09-266-965-62	Sequence 62, Appl
c	40	77.6	6.2	53500	4	US-09-266-965-76	Sequence 76, Appl
	41	75.2	6.0	1209	3	US-09-105-537-21	Sequence 21, Appl
	42	75.2	6.0	13613	3	US-09-105-537-3	Sequence 3, Appli
	43	75.2	6.0	38506	3	US-09-320-878-19	Sequence 19, Appl
	44	75.2	6.0	38506	4	US-09-141-908-1	Sequence 1, Appli
	45	75.2	6.0	38506	4	US-09-657-440-19	Sequence 19, Appl

ALIGNMENTS

```
RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-335-409-1

  Query Match             100.0%;  Score 1260;  DB 3;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-290;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728
```

```
Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628
```

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

```
  Query Match             100.0%;  Score 1260;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-290;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208
```

```
Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628


RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             100.0%;  Score 1260;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-290;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628


RESULT 4
US-09-568-480-1
```

```
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             100.0%;  Score 1260;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-290;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628


RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             100.0%;  Score 1260;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-290;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy          121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy          181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy          241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy          301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy          361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy          421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy          481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy          541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy          601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy          661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy          721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy          781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy          841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy          901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy          961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy         1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy         1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy         1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy         1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628
```

```
RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             100.0%;  Score 1260;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-290;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy            1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy           61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy          121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy          181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy          241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy          301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy          361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy          421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy          481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy          541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy         601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy         661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy         721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy         781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy         841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy         901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy         961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy        1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy        1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy        1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy        1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628

RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             100.0%;  Score 1260;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-290;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy          61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy         121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy         181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy         241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy         301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy         361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy         421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy         481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy         541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy         601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy         661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy         721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy         781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy         841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy         901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy         961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy        1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy         1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy         1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy         1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628
```

RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

```
  Query Match             98.9%;  Score 1246.2;  DB 4;  Length 71989;
  Best Local Similarity   99.4%;  Pred. No. 1.9e-286;
  Matches 1251;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy            1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        56757 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 56816

Qy           61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
                |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        56817 GCGCCTGGGTACGCGGAGGACCCGTTTCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 56876

Qy          121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        56877 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 56936

Qy          181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGAATCGAGCGCGGAGTAC 240
                |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        56937 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAATGGAATCGAGCGCGGAGTAC 56996

Qy          241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        56997 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 57056

Qy          301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
                ||||||||||||||||||||||||||||||||||| ||||||||||||||||| ||||||
Db        57057 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCATCGTTTACGTCACGCGCGATCGAC 57116

Qy          361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57117 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 57176

Qy          421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
                ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        57177 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGTGCGATCAGCGCT 57236

Qy          481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57237 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 57296

Qy          541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
                |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        57297 CGCGCGCTCGGCGTGGGTTTGGTGCCCCGGGTCGATGAGGAGACCAAGACCCTGGTCGCG 57356

Qy          601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
                |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        57357 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGGCGTCCTCGATGAGCGGCGCAGGAACCCG 57416

Qy          661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57417 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 57476

Qy          721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57477 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 57536

Qy          781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57537 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 57596

Qy          841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57597 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 57656

Qy          901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57657 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 57716

Qy          961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57717 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 57776

Qy         1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
                |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        57777 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGAGCGCGAGCCTCGCGTACGGTAGA 57836

Qy         1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57837 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 57896

Qy         1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57897 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 57956

Qy         1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATA 1259
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57957 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATA 58015
```

RESULT 9
US-09-252-991A-1265
; Sequence 1265, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

```
; TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE:  107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE:  1999-02-18
; PRIOR APPLICATION NUMBER:  US 60/074,788
; PRIOR FILING DATE:  1998-02-18
; PRIOR APPLICATION NUMBER:  US 60/094,190
; PRIOR FILING DATE:  1998-07-27
; NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 1265
;   LENGTH: 1314
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1265

  Query Match             11.5%;  Score 145.4;  DB 4;  Length 1314;
  Best Local Similarity   48.8%;  Pred. No. 1.1e-25;
  Matches  456;  Conservative    0;  Mismatches  446;  Indels   33;  Gaps    1;

Qy        254 CCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCGGAGGATCACGCTC 313
Db        290 CGGGGATCGGCGAGGGACTCAGCCACCACATGCTCAACCTCGACCCGCCGGACCATACCC 349

Qy        314 GGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGCGCGCCG 373
Db        350 GCCTGCGCTCGCTGGTCGGCCGCGCGTTCACCCCGCGCCAGGTGGAGCGCCTGCAACCGC 409

Qy        374 AAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACG 433
Db        410 ATATAGAACGGATCACCGAGGAATTGCTGGACGCCATGGCCGGCCGCGAACAGGCCGACC 469

Qy        434 TTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTC 493
Db        470 TGATGGCCGACTTCGCGATCCCGCTGACCATCGCGGTGATCTTCGAGCTGCTGGGCATTC 529

Qy        494 CGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCG 553
Db        530 CCGAGGCCGAGCGCGAACACGCCCGCCAGTCCTG------------------------- 563

Qy        554 TGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGG 613
Db        564 -------GGAGCGCCAGGCGGAACTGCTGTCGCCGGAGGAGGCCCAGGCCCTGGCCGATG 616

Qy        614 GGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACG 673
Db        617 CGCAGGTCGACTACCTGCGCGTGCTGCTCGAGGCCAAGCGCCGGCAGCCAGCCGACGACG 676

Qy        674 TCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGG 733
Db        677 TCTACAGCGGGCTGGTGCAGGCCGCCGACGAGAGCGGCCAACTGAGCGAGGCGGAACTCG 736

Qy        734 TCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCGT 793
Db        737 TCTCCATGGCCCACCTGCTGATGATGAGCGGCTTCGAGACCACCATGAACATGATCGGCA 796

Qy        794 TCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCG 853
Db        797 ACGCGCTGGTCACCCTGCTGGTCAACCCGGAGCAACTGGCGTTGCTGCGGGCGCAGCCGG 856

Qy        854 GGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAATAGGAA 913
Db        857 AACTCCTGCCCAACGCCATGGAAGAACTGGTCCGCCACGACAGCCCGGTGCGCGCCTCGA 916

Qy        914 CTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGGAGA 973
Db        917 TGTTGCGCTTCACCGTGGAAGACGTGGAACTGGACGGGGTCACCATTCCCGCCGGCGAAT 976

Qy        974 TGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTCTCCAGGCCAGACG 1033
Db        977 ACATCCTGGTCTCCAACCTGACCGCCAACCACGACGCCGAGCGCTTCGACGATCCCGACC 1036

Qy       1034 TGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCCCCATGTCT 1093
Db       1037 GCCTCGACCTCACCCGCAACACCGATGGCCATCTCGGCTACGGCTTCGGCGTGCACTACT 1096

Qy       1094 GCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCACCATCTTCCGTA 1153
Db       1097 GCGTCGGCGCCTCGCTGGCCCGGCTGGAGGGGCGGATCGCCATCCAGCGCCTGCTCGCGC 1156

Qy       1154 GGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTG 1188
Db       1157 GCTTCCCCGACCTCCAGTTGGCGGTGCCCCACGCG 1191

RESULT 10
US-09-252-991A-1338/c
; Sequence 1338, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 1338
;   LENGTH: 1482
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1338

  Query Match             11.5%;  Score 145.4;  DB 4;  Length 1482;
  Best Local Similarity   48.8%;  Pred. No. 1.2e-25;
  Matches  456;  Conservative    0;  Mismatches  446;  Indels   33;  Gaps    1;

Qy        254 CCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCGGAGGATCACGCTC 313
Db       1217 CGGGGATCGGCGAGGGACTCAGCCACCACATGCTCAACCTCGACCCGCCGGACCATACCC 1158

Qy        314 GGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGCGCGCCG 373
Db       1157 GCCTGCGCTCGCTGGTCGGCCGCGCGTTCACCCCGCGCCAGGTGGAGCGCCTGCAACCGC 1098

Qy        374 AAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACG 433
Db       1097 ATATAGAACGGATCACCGAGGAATTGCTGGACGCCATGGCCGGCCGCGAACAGGCCGACC 1038

Qy        434 TTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTC 493
Db       1037 TGATGGCCGACTTCGCGATCCCGCTGACCATCGCGGTGATCTTCGAGCTGCTGGGCATTC 978

Qy        494 CGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCG 553
Db        977 CCGAGGCCGAGCGCGAACACGCCCGCCAGTCCTG------------------------- 944

Qy        554 TGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGG 613
Db        943 -------GGAGCGCCAGGCGGAACTGCTGTCGCCGGAGGAGGCCCAGGCCCTGGCCGATG 891

Qy        614 GGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACG 673
Db        890 CGCAGGTCGACTACCTGCGCGTGCTGCTCGAGGCCAAGCGCCGGCAGCCAGCCGACGACG 831

Qy        674 TCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGG 733
Db        830 TCTACAGCGGGCTGGTGCAGGCCGCCGACGAGAGCGGCCAACTGAGCGAGGCGGAACTCG 771

Qy        734 TCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCGT 793
Db        770 TCTCCATGGCCCACCTGCTGATGATGAGCGGCTTCGAGACCACCATGAACATGATCGGCA 711
```

```
Qy        794 TCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCG 853
Db        710 ACGCGCTGGTCACCCTGCTGGTCAACCCGGAGCAACTGGCGTTGCTGCGGGCGCAGCCGG 651

Qy        854 GGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAATAGGAA 913
Db        650 AACTCCTGCCCAACGCCATGGAAGAACTGGTCCGCCACGACAGCCCGGTGCGCGCCTCGA 591

Qy        914 CTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGGAGA 973
Db        590 TGTTGCGCTTCACCGTGGAAGACGTGGAACTGGACGGGGTCACCATTCCCGCCGGCGAAT 531

Qy        974 TGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTCTCCAGGCCAGACG 1033
Db        530 ACATCCTGGTCTCCAACCTGACCGCCAACCACGACGCCGAGCGCTTCGACGATCCCGACC 471

Qy       1034 TGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCCCCATGTCT 1093
Db        470 GCCTCGACCTCACCCGCAACACCGATGGCCATCTCGGCTACGGCTTCGGCGTGCACTACT 411

Qy       1094 GCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGCACCATCTTCCGTA 1153
Db        410 GCGTCGGCGCCTCGCTGGCCCGGCTGGAGGGGCGGATCGCCATCCAGCGCCTGCTCGCGC 351

Qy       1154 GGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTG 1188
Db        350 GCTTCCCCGACCTCCAGTTGGCGGTGCCCCACGCG 316
```

RESULT 11
US-09-252-991A-1222
; Sequence 1222, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1222
; LENGTH: 1671
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1222

```
  Query Match             11.5%;  Score 145.4;  DB 4;  Length 1671;
  Best Local Similarity   48.8%;  Pred. No. 1.2e-25;
  Matches  456;  Conservative    0;  Mismatches  446;  Indels   33;  Gaps    1;

Qy        254 CCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCGGAGGATCACGCTC 313
Db        225 CGGGGATCGGCGAGGGACTCAGCCACCACATGCTCAACCTCGACCCGCCGGACCATACCC 284

Qy        314 GGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGCGCGCCG 373
Db        285 GCCTGCGCTCGCTGGTCGGCCGCGCGTTCACCCCGCGCCAGGTGGAGCGCCTGCAACCGC 344

Qy        374 AAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACG 433
Db        345 ATATAGAACGGATCACCGAGGAATTGCTGGACGCCATGGCCGGCCGCGAACAGGCCGACC 404

Qy        434 TTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTC 493
Db        405 TGATGGCCGACTTCGCGATCCCGCTGACCATCGCGGTGATCTTCGAGCTGCTGGGCATTC 464

Qy        494 CGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCG 553
Db        465 CCGAGGCCGAGCGCGAACACGCCCGCCAGTCCTG------------------------ 498

Qy        554 TGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGG 613
Db        499 -------GGAGCGCCAGGCGGAACTGCTGTCGCCGGAGGAGGCCCAGGCCCTGGCCGATG 551

Qy        614 GGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACG 673
Db        552 CGCAGGTCGACTACCTGCGCGTGCTGCTCGAGGCCAAGCGCCGGCAGCCAGCCGACGACG 611

Qy        674 TCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGG 733
Db        612 TCTACAGCGGGCTGGTGCAGGCCGCCGACGAGAGCGGCCAACTGAGCGAGGCGGAACTCG 671

Qy        734 TCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCGT 793
Db        672 TCTCCATGGCCCACCTGCTGATGATGAGCGGCTTCGAGACCACCATGAACATGATCGGCA 731

Qy        794 TCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCG 853
Db        732 ACGCGCTGGTCACCCTGCTGGTCAACCCGGAGCAACTGGCGTTGCTGCGGGCGCAGCCGG 791

Qy        854 GGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAATAGGAA 913
Db        792 AACTCCTGCCCAACGCCATGGAAGAACTGGTCCGCCACGACAGCCCGGTGCGCGCCTCGA 851

Qy        914 CTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGGAGA 973
Db        852 TGTTGCGCTTCACCGTGGAAGACGTGGAACTGGACGGGGTCACCATTCCCGCCGGCGAAT 911

Qy        974 TGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTCTCCAGGCCAGACG 1033
Db        912 ACATCCTGGTCTCCAACCTGACCGCCAACCACGACGCCGAGCGCTTCGACGATCCCGACC 971

Qy       1034 TGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCCCCATGTCT 1093
Db        972 GCCTCGACCTCACCCGCAACACCGATGGCCATCTCGGCTACGGCTTCGGCGTGCACTACT 1031

Qy       1094 GCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGCACCATCTTCCGTA 1153
Db       1032 GCGTCGGCGCCTCGCTGGCCCGGCTGGAGGGGCGGATCGCCATCCAGCGCCTGCTCGCGC 1091

Qy       1154 GGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTG 1188
Db       1092 GCTTCCCCGACCTCCAGTTGGCGGTGCCCCACGCG 1126
```

RESULT 12
US-09-105-537-38
; Sequence 38, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-38

```
  Query Match             9.8%;  Score 123;  DB 3;  Length 1251;
  Best Local Similarity   47.5%;  Pred. No. 2.3e-20;
  Matches  452;  Conservative    0;  Mismatches  475;  Indels   24;  Gaps    2;

Qy        236 AGTACTCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGC 295
```

```
Db        227 ACTCCACGACTCCCCTGACCGAGGCCGAGGCCGCGCTCAACCACAACATGCTGGAGTCCG 286

Qy        296 CGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCA 355

Db        287 ACCCGCCGCGGCACACCCGGCTGCGCAAGCTGGTGGCCCGTGAGTTCACCATGCGCCGGG 346

Qy        356 TCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCG 415

Db        347 TCGAGTTGCTGCGGCCCCGGGTCCAGGAGATCGTCGACGGGCTCGTGGACGCCATGCTGG 406

Qy        416 GACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCA 475

Db        407 CGGCGCCCGACGGCCGCGCCGATCTGATGGAGTCCCTGGCCTGGCCGCTGC---CGATCA 463

Qy        476 GCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGA 535

Db        464 CCGTGATCTCCGAACTCCTCGGCGTGCCCGAGCCGGACCGCGCCGCCTTCCGCGTCTGGA 523

Qy        536 CTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGG 595

Db        524 CCGACGCCTTCGTCTTCCCGGACGATCCCGCCCAGGCCCAGACCG--------------- 568

Qy        596 TCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGA 655

Db        569 ------CCATGGCCGAGATGAGCGGCTATCTCTCCCGGCTCATCGACTCCAAGCGCGGGC 622

Qy        656 ACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGC 715

Db        623 AGGACGGCGAGGACCTGCTCAGCGCGCTCGTGCGGACCAGCGACGAGGACGGCTCCCGGC 682

Qy        716 TGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCA 775

Db        683 TGACCTCCGAGGAGCTGCTCGGTATGGCCCACATCCTGCTCGTCGCGGGGCACGAGACCA 742

Qy        776 CGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGC 835

Db        743 CGGTCAATCTGATCGCCAACGGCATGTACGCGCTGCTCTCGCACCCCGACCAGCTGGCCG 802

Qy        836 TGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACA 895

Db        803 CCCTGCGGGCCGACATGACGCTCTTGGACGGCGCGGTGGAGGAGATGTTGCGCTACGAGG 862

Qy        896 ATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCAT 955

Db        863 GCCCGGTGGAATCCGCGACCTACCGCTTCCCGGTCGAGCCCGTCGACCTGGACGGCACGG 922

Qy        956 CGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGACTG 1015

Db        923 TCATCCCGGCCGGTGACACGGTCCTCGTCGTCCTGGCCGACGCCCACCGCACCCCCGAGC 982

Qy       1016 TATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACG 1075

Db        983 GCTTCCCGGACCCGCACCGCTTCGACATCCGCCGGGACACCGCCGGCCATCTCGCCTTCG 1042

Qy       1076 GTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCG 1135

Db       1043 GCCACGGCATCCACTTCTGCATCGGCGCCCCCTTGGCCCGGTTGGAGGCCCGGATCGCCG 1102

Qy       1136 TGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCG 1186

Db       1103 TCCGCGCCCTTCTCGAACGCTGCCCGGACCTCGCCCTGGACGTCTCCCCCG 1153
```

RESULT 13
US-09-320-878-21
; Sequence 21, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-21

```
  Query Match             9.8%;  Score 123;  DB 3;  Length 5970;
  Best Local Similarity   47.5%;  Pred. No. 3.3e-20;
  Matches  452;  Conservative    0;  Mismatches  475;  Indels   24;  Gaps    2;

Qy        236 AGTACTCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGC 295

Db       1582 ACTCCACGACTCCCCTGACCGAGGCCGAGGCCGCGCTCAACCACAACATGCTGGAGTCCG 1641

Qy        296 CGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCA 355

Db       1642 ACCCGCCGCGGCACACCCGGCTGCGCAAGCTGGTGGCCCGTGAGTTCACCATGCGCCGGG 1701

Qy        356 TCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCG 415

Db       1702 TCGAGTTGCTGCGGCCCCGGGTCCAGGAGATCGTCGACGGGCTCGTGGACGCCATGCTGG 1761

Qy        416 GACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCA 475

Db       1762 CGGCGCCCGACGGCCGCGCCGATCTGATGGAGTCCCTGGCCTGGCCGCTGC---CGATCA 1818

Qy        476 GCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGA 535

Db       1819 CCGTGATCTCCGAACTCCTCGGCGTGCCCGAGCCGGACCGCGCCGCCTTCCGCGTCTGGA 1878

Qy        536 CTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGG 595

Db       1879 CCGACGCCTTCGTCTTCCCGGACGATCCCGCCCAGGCCCAGACCG--------------- 1923

Qy        596 TCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGA 655

Db       1924 ------CCATGGCCGAGATGAGCGGCTATCTCTCCCGGCTCATCGACTCCAAGCGCGGGC 1977

Qy        656 ACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGC 715

Db       1978 AGGACGGCGAGGACCTGCTCAGCGCGCTCGTGCGGACCAGCGACGAGGACGGCTCCCGGC 2037

Qy        716 TGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCA 775

Db       2038 TGACCTCCGAGGAGCTGCTCGGTATGGCCCACATCCTGCTCGTCGCGGGGCACGAGACCA 2097

Qy        776 CGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGC 835

Db       2098 CGGTCAATCTGATCGCCAACGGCATGTACGCGCTGCTCTCGCACCCCGACCAGCTGGCCG 2157

Qy        836 TGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACA 895

Db       2158 CCCTGCGGGCCGACATGACGCTCTTGGACGGCGCGGTGGAGGAGATGTTGCGCTACGAGG 2217
```

```
Qy        896 ATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCAT 955
Db       2218 GCCCGGTGGAATCCGCGACCTACCGCTTCCCGGTCGAGCCCGTCGACCTGGACGGCACGG 2277

Qy        956 CGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTG 1015
Db       2278 TCATCCCGGCCGGTGACACGGTCCTCGTCGTCCTGGCCGACGCCCACCGCACCCCCGAGC 2337

Qy       1016 TATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACG 1075
Db       2338 GCTTCCCGGACCCGCACCGCTTCGACATCCGCCGGGACACCGCCGGCCATCTCGCCTTCG 2397

Qy       1076 GTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCG 1135
Db       2398 GCCACGGCATCCACTTCTGCATCGGCGCCCCCTTGGCCCGGTTGGAGGCCCGGATCGCCG 2457

Qy       1136 TGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCG 1186
Db       2458 TCCGCGCCCTTCTCGAACGCTGCCCGGACCTCGCCCTGGACGTCTCCCCCG 2508
```

RESULT 14
US-09-141-908-11
; Sequence 11, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-11

```
  Query Match             9.8%;  Score 123;  DB 4;  Length 5970;
  Best Local Similarity   47.5%;  Pred. No. 3.3e-20;
  Matches  452;  Conservative    0;  Mismatches  475;  Indels   24;  Gaps    2;

Qy        236 AGTACTCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGC 295
Db       1582 ACTCCACGACTCCCCTGACCGAGGCCGAGGCCGCGCTCAACCACAACATGCTGGAGTCCG 1641

Qy        296 CGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCA 355
Db       1642 ACCCGCCGCGGCACACCCGGCTGCGCAAGCTGGTGGCCCGTGAGTTCACCATGCGCCGGG 1701

Qy        356 TCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCG 415
Db       1702 TCGAGTTGCTGCGGCCCCGGGTCCAGGAGATCGTCGACGGGCTCGTGGACGCCATGCTGG 1761

Qy        416 GACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCA 475
Db       1762 CGGCGCCCGACGGCCGCGCCGATCTGATGGAGTCCCTGGCCTGGCCGCTGC---CGATCA 1818

Qy        476 GCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGA 535
Db       1819 CCGTGATCTCCGAACTCCTCGGCGTGCCCGAGCCGGACCGCGCCGCCTTCCGCGTCTGGA 1878

Qy        536 CTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGG 595
Db       1879 CCGACGCCTTCGTCTTCCCGGACGATCCCGCCCAGGCCCAGACCG-------------- 1923

Qy        596 TCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGA 655
Db       1924 ------CCATGGCCGAGATGAGCGGCTATCTCTCCCGGCTCATCGACTCCAAGCGCGGGC 1977

Qy        656 ACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGC 715
Db       1978 AGGACGGCGAGGACCTGCTCAGCGCGCTCGTGCGGACCAGCGACGAGGACGGCTCCCGGC 2037

Qy        716 TGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCA 775
Db       2038 TGACCTCCGAGGAGCTGCTCGGTATGGCCCACATCCTGCTCGTCGCGGGGCACGAGACCA 2097

Qy        776 CGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGC 835
Db       2098 CGGTCAATCTGATCGCCAACGGCATGTACGCGCTGCTCTCGCACCCCGACCAGCTGGCCG 2157

Qy        836 TGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACA 895
Db       2158 CCCTGCGGGCCGACATGACGCTCTTGGACGGCGCGGTGGAGGAGATGTTGCGCTACGAGG 2217

Qy        896 ATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCAT 955
Db       2218 GCCCGGTGGAATCCGCGACCTACCGCTTCCCGGTCGAGCCCGTCGACCTGGACGGCACGG 2277

Qy        956 CGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTG 1015
Db       2278 TCATCCCGGCCGGTGACACGGTCCTCGTCGTCCTGGCCGACGCCCACCGCACCCCCGAGC 2337

Qy       1016 TATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACG 1075
Db       2338 GCTTCCCGGACCCGCACCGCTTCGACATCCGCCGGGACACCGCCGGCCATCTCGCCTTCG 2397

Qy       1076 GTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCG 1135
Db       2398 GCCACGGCATCCACTTCTGCATCGGCGCCCCCTTGGCCCGGTTGGAGGCCCGGATCGCCG 2457

Qy       1136 TGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCG 1186
Db       2458 TCCGCGCCCTTCTCGAACGCTGCCCGGACCTCGCCCTGGACGTCTCCCCCG 2508
```

RESULT 15
US-09-657-440-21
; Sequence 21, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-657-440-21

```
Query Match             9.8%;  Score 123;  DB 4;  Length 5970;
Best Local Similarity   47.5%;  Pred. No. 3.3e-20;
Matches  452;  Conservative    0;  Mismatches  475;  Indels   24;  Gaps    2;

Qy        236 AGTACTCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGC 295

Db       1582 ACTCCACGACTCCCCTGACCGAGGCCGAGGCCGCGCTCAACCACAACATGCTGGAGTCCG 1641

Qy        296 CGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCA 355

Db       1642 ACCCGCCGCGGCACACCCGGCTGCGCAAGCTGGTGGCCCGTGAGTTCACCATGCGCCGGG 1701

Qy        356 TCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCG 415

Db       1702 TCGAGTTGCTGCGGCCCCGGGTCCAGGAGATCGTCGACGGGCTCGTGGACGCCATGCTGG 1761

Qy        416 GACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCA 475

Db       1762 CGGCGCCCGACGGCCGCGCCGATCTGATGGAGTCCCTGGCCTGGCCGCTGC---CGATCA 1818

Qy        476 GCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGA 535

Db       1819 CCGTGATCTCCGAACTCCTCGGCGTGCCCGAGCCGGACCGCGCCGCCTTCCGCGTCTGGA 1878

Qy        536 CTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGG 595

Db       1879 CCGACGCCTTCGTCTTCCCGGACGATCCCGCCCAGGCCCAGACCG-------------- 1923

Qy        596 TCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGA 655

Db       1924 ------CCATGGCCGAGATGAGCGGCTATCTCTCCCGGCTCATCGACTCCAAGCGCGGGC 1977

Qy        656 ACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGC 715

Db       1978 AGGACGGCGAGGACCTGCTCAGCGCGCTCGTGCGGACCAGCGACGAGGACGGCTCCCGGC 2037

Qy        716 TGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCA 775

Db       2038 TGACCTCCGAGGAGCTGCTCGGTATGGCCCACATCCTGCTCGTCGCGGGGCACGAGACCA 2097

Qy        776 CGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGC 835

Db       2098 CGGTCAATCTGATCGCCAACGGCATGTACGCGCTGCTCTCGCACCCCGACCAGCTGGCCG 2157

Qy        836 TGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACA 895

Db       2158 CCCTGCGGGCCGACATGACGCTCTTGGACGGCGCGGTGGAGGAGATGTTGCGCTACGAGG 2217

Qy        896 ATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCAT 955

Db       2218 GCCCGGTGGAATCCGCGACCTACCGCTTCCCGGTCGAGCCCGTCGACCTGGACGGCACGG 2277

Qy        956 CGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTG 1015

Db       2278 TCATCCCGGCCGGTGACACGGTCCTCGTCGTCCTGGCCGACGCCCACCGCACCCCCGAGC 2337

Qy       1016 TATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACG 1075

Db       2338 GCTTCCCGGACCCGCACCGCTTCGACATCCGCCGGGACACCGCCGGCCATCTCGCCTTCG 2397

Qy       1076 GTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCG 1135

Db       2398 GCCACGGCATCCACTTCTGCATCGGCGCCCCCTTGGCCCGGTTGGAGGCCCGGATCGCCG 2457

Qy       1136 TGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCG 1186

Db       2458 TCCGCGCCCTTCTCGAACGCTGCCCGGACCTCGCCCTGGACGTCTCCCCCG 2508

Search completed: October  4, 2003, 03:40:32
Job time : 83.7507 secs
```

OM nucleic - nucleic search, using sw model

```
Run on:          October  3, 2003, 11:29:26 ; Search time 308.318 Seconds
                                          (without alignments)
                                          11031.749 Million cell updates/sec

Title:           US-10-014-717-1_COPY_62369_63628
Perfect score:   1260
Sequence:        1 atgacacaggagcaagcgaa..........agccctccaaagctggatag 1260

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters:      5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_19Jun03:*
               1:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
               2:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
               3:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
               4:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
               5:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
               6:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
               7:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
               8:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
               9:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
               10:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
               11:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
               12:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
               13:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
               14:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
               15:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
               16:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
               17:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
               18:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
               19:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
               20:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
               21:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
               22:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
               23:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
               24:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
               25:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1258.4	99.9	68750	21	AAZ55887	Sorangium cellulos
	2	1246.2	98.9	71989	21	AAA29349	Sorangium cellulos
c	3	149.2	11.8	67459	25	ABQ77491	S. aurantiaca DNA
	4	143.8	11.4	1257	22	AAF81357	Quorum sensing con
	5	131.6	10.4	1400	12	AAQ14548	EryF gene. Saccar
	6	128.8	10.2	1347	25	ABX56047	M. echinospora cal
	7	123	9.8	5970	21	AAA75635	Nucleotide sequenc
	8	123	9.8	5970	21	AAZ56003	Contig 002 from co
	9	123	9.8	5970	25	ABS56092	Contig 002 from co
	10	122.6	9.7	1191	24	ABN88910	Streptomyces sp. c
	11	121.4	9.6	1251	21	AAZ87301	S. venezuelae macr
	12	120.8	9.6	1992	25	ABZ76672	Streptomyces sp. T
c	13	119.6	9.5	36538	25	ABV75558	Saccharopolyspora
	14	118.6	9.4	1449	24	ABN88911	Streptomyces sp. c
	15	116.8	9.3	1209	25	ABX56043	M. echinospora cal
	16	115	9.1	1879	12	AAQ11126	Sequence encoding
	17	114.8	9.1	53789	19	AAV21187	Amycolatopsis medi
	18	114.6	9.1	1428	25	AAD53024	Streptomyces albof
	19	114.4	9.1	1227	22	AAD14511	Streptomyces clavu
	20	114.4	9.1	1233	21	AAC55790	Hydroxylase orf3 i
	21	114.4	9.1	12249	21	AAC55840	Complete Mitomycin
	22	114.4	9.1	15079	22	AAD14499	Streptomyces clavu
	23	114.4	9.1	18331	21	AAC55857	Complete nucleotid
	24	112.8	9.0	15079	16	AAQ91580	S. clavuligerus cl
	25	111.8	8.9	1293	25	AAD53020	Streptomyces lydic
	26	111.8	8.9	1293	25	AAD53026	Streptomyces kasug
	27	111.4	8.8	1428	25	AAD53027	Streptomyces rimos
	28	111.4	8.8	1700	14	AAQ45569	Sequence soyC and
	29	110	8.7	966	24	ABK74899	Bacillus lichenifo
	30	109.6	8.7	1224	21	AAC55788	Hydroxylase orf4 i
	31	109.4	8.7	103599	23	ABX04971	S. cinnamonensis m
	32	108.6	8.6	1293	25	AAD53031	Streptomyces lydic
	33	107.8	8.6	135638	25	ABX34289	S. atroolivaceus 1
	34	107	8.5	1293	25	AAD53022	Streptomyces chatt
	35	107	8.5	2795	22	AAI71296	Streptomyces grise
	36	107	8.5	2795	22	AAF81954	Streptomyces grise
	37	106.6	8.5	1413	25	AAD53019	Streptomyces rimos
	38	106.4	8.4	2918	21	AAZ43922	S. tendae nikkomyc
	39	106.4	8.4	2919	21	AAZ24486	Streptomyces tende
	40	98	7.8	1293	25	AAD53052	Streptomyces tuber
	41	97.4	7.7	50937	21	AAA09469	Streptococcus olea
	42	95.4	7.6	1293	25	AAD53028	Streptomyces tuber
	43	94.8	7.5	1998	12	AAQ11127	Sequence encoding
c	44	94.6	7.5	47981	22	AAF30757	Micromonospora meg
	45	94	7.5	1293	25	AAD53032	Streptomyces lydic

ALIGNMENTS

```
RESULT 1
AAZ55887
ID   AAZ55887 standard; DNA; 68750 BP.
XX
AC   AAZ55887;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum 68.75 kb contig.
XX
KW   Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW   anticancer; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1826
FT                   /*tag=  a
FT                   /partial
FT                   /product= "Partial Orf 1 protein (AAY58580)"
FT                   /note= "No initiation codon given in the specification"
FT   CDS             complement (1900..3171)
FT                   /*tag=  b
FT                   /product= "Orf 2 protein (AAY58581)"
FT   CDS             3415..5556
FT                   /*tag=  c
FT                   /product= "Orf 3 protein (AAY58582)"
FT   CDS             complement (5612..5992)
FT                   /*tag=  d
FT                   /product= "Orf 4 protein (AAY58583)"
FT   CDS             6226..6675
```

```
FT                   /*tag= e
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                             (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                             (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                             (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                             (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                             (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                             (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                             (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR   AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR   AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR   AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy  -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             99.9%;  Score 1258.4;  DB 21;  Length 68750;
  Best Local Similarity   99.9%;  Pred. No. 3.1e-279;
  Matches 1259;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy         421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy         481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy         541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy         601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy         661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy         721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy         781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy         841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy         901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy         961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
               ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       63329 AAGAAAGGGGAGATGGTCTTTCTCCTCATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy        1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy        1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy        1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy        1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628
```

```
RESULT 2
AAA29349
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..992
FT                   /*tag=  a
FT                   /label= ORF_A
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             989..1501
FT                   /*tag=  b
FT                   /label= ORF_B
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             1998..6263
FT                   /*tag=  c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031..3548
FT                   /*tag=  d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                          domain"
FT   misc_RNA        3621..4661
FT                   /*tag=  e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                          domain"
FT   misc_RNA        4917..5810
FT                   /*tag=  f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                          domain, potentially involved in formation of the
FT                          thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag=  g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                          domain"
FT   CDS             6260..10493
FT                   /*tag=  h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag=  i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                          module"
FT   misc_RNA        2031..3548
FT                   /*tag=  j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                          module"
FT   misc_RNA        6861..6887
FT                   /*tag=  k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag=  l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                          module"
FT   misc_RNA        7358..7366
FT                   /*tag=  m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                          NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag=  n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag=  o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag=  p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag=  q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag=  r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
```

```
FT   misc_RNA        8966..8983
FT                   /*tag=  s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag=  t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag=  u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag=  v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                          module"
FT   misc_RNA        10261..10306
FT                   /*tag=  w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                          module"
FT   CDS             10639..16137
FT                   /*tag=  x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag=  y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag=  z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag=  aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                          module 2"
FT   misc_RNA        14962..15756
FT                   /*tag=  ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                          module 2"
FT   misc_RNA        15763..16008
FT                   /*tag=  ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag=  ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag=  ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag=  af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag=  ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag=  ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag=  ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag=  aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag=  ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag=  al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag=  am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag=  an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag=  ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag=  ap
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag=  aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag=  ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag=  as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag=  at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag=  au
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag=  av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag=  aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag=  ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag=  ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag=  az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag=  ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag=  bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             98.9%;  Score 1246.2;  DB 21;  Length 71989;
  Best Local Similarity   99.4%;  Pred. No. 2e-276;
  Matches 1251;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56757 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 56816

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db      56817 GCGCCTGGGTACGCGGAGGACCCGTTTCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 56876

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56877 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 56936

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGAATCGAGCGCGGAGTAC 240
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      56937 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAATGGAATCGAGCGCGGAGTAC 56996

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56997 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 57056

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||| ||||||
Db      57057 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCATCGTTTACGTCACGCGCGATCGAC 57116

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57117 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 57176

Qy          421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
                ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        57177 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGTGCGATCAGCGCT 57236

Qy          481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57237 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 57296

Qy          541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
                |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        57297 CGCGCGCTCGGCGTGGGTTTGGTGCCCCGGGTCGATGAGGAGACCAAGACCCTGGTCGCG 57356

Qy          601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
                |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        57357 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGGCGTCCTCGATGAGCGGCGCAGGAACCCG 57416

Qy          661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57417 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 57476

Qy          721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57477 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 57536

Qy          781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57537 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 57596

Qy          841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57597 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 57656

Qy          901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57657 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 57716

Qy          961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57717 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 57776

Qy         1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
                |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        57777 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGAGCGCGAGCCTCGCGTACGGTAGA 57836

Qy         1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57837 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 57896

Qy         1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57897 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 57956

Qy         1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATA 1259
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        57957 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATA 58015
```

```
RESULT 3
ABQ77491/c
ID   ABQ77491 standard; DNA; 67459 BP.
XX
AC   ABQ77491;
XX
DT   14-MAY-2003  (first entry)
XX
DE   S. aurantiaca DNA containing sti gene cluster.
XX
KW   stiA; stiB; stiC; stiD; stiE; stiF; stiG; stiH; stiJ; stiK; stiL;
KW   aromatic structure biosynthesis; beta-ketoacylsynthase;
KW   acyltransferase; dehydratase; enoylreductase; beta-ketoacylreductase;
KW   acyl carrier protein; O-methyl transferase; ds.
XX
OS   Stigmatella aurantiaca.
XX
FH   Key             Location/Qualifiers
FT   CDS             14345..21466
FT                   /*tag=  a
FT                   /product= "stiA"
FT                   /note= "GTG start codon"
FT   CDS             21463..26220
FT                   /*tag=  b
FT                   /product= "stiB"
FT   CDS             26284..31941
FT                   /*tag=  c
FT                   /product= "stiC"
FT   CDS             31934..37738
FT                   /*tag=  d
FT                   /product= "stiD"
FT   CDS             37740..43553
FT                   /*tag=  e
FT                   /product= "stiE"
FT   CDS             43546..50202
FT                   /*tag=  f
FT                   /product= "stiF"
FT   CDS             50205..54401
FT                   /*tag=  g
FT                   /product= "stiG"
FT   CDS             54398..59212
FT                   /*tag=  h
FT                   /product= "stiH"
FT   CDS             59223..63002
FT                   /*tag=  i
FT                   /product= "stiJ"
FT   CDS             63070..63840
FT                   /*tag=  j
FT                   /product= "stiK"
FT   CDS             63929..65470
FT                   /*tag=  k
FT                   /product= "stiL"
FT                   /note= "GTG start codon"
XX
PN   DE10128661-A1.
XX
PD   19-DEC-2002.
XX
PF   15-JUN-2001; 2001DE-1028661.
XX
PR   15-JUN-2001; 2001DE-1028661.
XX
PA   (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX
PI   Kunze B,  Nordsiek G,  Silakowski B,  Bloecker H,  Gaitatzis N;
PI   Mueller R;
XX
DR   WPI; 2003-185404/19.
XX
PT   Nucleic acid molecule comprising one or more genes from Stigmatella
PT   species -
XX
PS   Claim 3; Page 9-43; 48pp; German.
XX
CC   This invention describes a novel nucleic acid molecule comprising one or
CC   more genes from Stigmatella species selected from stiA, stiB, stiC, stiD,
CC   stiE, stiF, stiG, stiH, stiJ stiK and stiL. The nucleic acid is used for
CC   biosynthetic purposes, particularly for producing aromatic structures.
CC   The products of the invention have domains sharing high homology with
CC   beta-ketoacylsynthase (KS), acyltransferase (AT), dehydratase (DH),
CC   enoylreductase (ER), beta-ketoacylreductase (KR), acyl carrier protein
CC   (ACP), O-methyl transferase (O-MT). This sequence represents a
CC   polynucleotide isolated from S. aurantiaca which contains the sti gene
CC   cluster described in the disclosure of the invention.
XX
SQ   Sequence 67459 BP; 10755 A; 20921 C; 23348 G; 12435 T; 0 other;
```

```
  Query Match             11.8%;  Score 149.2;  DB 25;  Length 67459;
  Best Local Similarity   49.2%;  Pred. No. 1.7e-24;
  Matches  450;  Conservative    0;  Mismatches  458;  Indels    6;  Gaps    2;

Qy        298 CCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATC 357
Db       5928 CCGCCGCTGCACACGCGGGTGCGCCGACTGATCGTGGGCGCGCTGACGCCGCGCGCGTTG 5869

Qy        358 GACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGA 417
Db       5868 TCGACAATGGAGCCAGGGTTGCGCACCCTGGTGGACCGGCTGCTGGACGGGCTGGCCGTG 5809

Qy        418 CAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGC 477
Db       5808 AAGGGCGCCGCAGACCTGATCGAGGACTTCGCCGCCGCCATCCCCATCGAGGTAATTGGC 5749

Qy        478 GCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACT 537
Db       5748 AACCTGCTGGACATCCCGACGGAGGAGCGGGGGCCGCTGCGGGGCTGGTCGCTCGCGATC 5689

Qy        538 GCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTC 597
Db       5688 CTCGGCGCGCTCGAACCCCGTCTGACGGCCGAGCAGGAGGCCTGCGGGAATGAGGCGGTA 5629

Qy        598 GCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAAC 657
Db       5628 ACAGAGTTCCTGGACTACCTCCGCATCCTCGTCGCGCAGCGGCGAGCGCGGCCGGGGGAC 5569

Qy        658 CCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTG 717
Db       5568 CC---CGCAACGGACGTGCTGACGCGCCTCATCCAGGGCGAATCCGACGGGGAACGATTG 5512

Qy        718 AGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACG 777
Db       5511 ACCGAGACGGAGCTGCTGCACCAGTGCGTCTTCCTCCTGAACGCGGGGCACGAGACGACA 5452

Qy        778 ATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTG 837
Db       5451 ACGAACCTCATCGGCAACGCCCTGGAACTGCTGGCGCGCTTCCCGGACGAGCGCGCTCGG 5392

Qy        838 GTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAAT 897
Db       5391 CTGCTGCGAGCCCCCGCGCTGATCCCGACGGCAGTGGAGGAGGTCCTGCGTTACGAATCC 5332

Qy        898 ATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCG 957
Db       5331 TCCAACCAGCTCGGCAACCGGCGGGTGGC---GGAGGACACGGAGATTGGCGGCGTAGCG 5275

Qy        958 ATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTA 1017
Db       5274 GTGCCGACCGGCACCTTCCTCACGCTCTGCATCGGCGCCGCCAACCGCGACCCCGCGCGC 5215

Qy       1018 TTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGT 1077
Db       5214 TTCGAGGACCCAGAGCACTTCGACGTGGGGCGCCAGCCCAACCGGCACCTCGCCTTCGCG 5155

Qy       1078 AGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTG 1137
Db       5154 GGCGGGGCGCATACGTGCGCGGGGATGAACCTTGCGCGGATAGAGGCGCGGATAGCGTTG 5095

Qy       1138 GGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATAC 1197
Db       5094 GCGGCGTTCCTCGCGCGCTTCCCGGACTATGCGCTGACCGCACCGCCGGTGAGGGCGCGG 5035

Qy       1198 CACCCCGCGTTCCG 1211
Db       5034 CGGGCGCGGTTCCG 5021


RESULT 4
AAF81357
ID   AAF81357 standard; DNA; 1257 BP.
XX
AC   AAF81357;
XX
DT   04-JUN-2001  (first entry)
XX
DE   Quorum sensing controlled gene qsc122 ORF.
XX
KW   Quorum sensing; antibacterial; bacterial signalling;
KW   opportunistic pathogen; immunocompromised; burn; cystic fibrosis;
KW   immunosuppressive therapy; AIDS; ss.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200118248-A2.
XX
PD   15-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-US24141.
XX
PR   03-SEP-1999;   99US-0153022.
XX
PA   (IOWA ) UNIV IOWA RES FOUND.
PA   (QUOR-) QUORUM SCI INC.
XX
PI   Whiteley M,  Lee KM,  Greenberg EP,  Muh U;
XX
DR   WPI; 2001-265973/27.
XX
PT   Identifying modulators of quorum sensing signaling in Pseudomonas
PT   aeruginosa bacteria, useful for treating infections in
PT   immunocompromized patients -
XX
PS   Claim 46; Page 101-102; 115pp; English.
XX
CC   Bacteria signal to one another to coordinate expression of specific genes
CC   in a cell density dependent fashion. This "bacterial signalling" is
CC   called "quorum sensing and response". Quorum sensing allows a bacterial
CC   species to sense its own number and regulate gene expression according to
CC   population density. The present sequence is an open reading frame (ORF)
CC   of a Pseudomonas aeruginosa quorum sensing controlled gene. Inhibitors of
CC   quorum sensing signalling renders a bacterial population more susceptible
CC   to treatment. The present invention relates to a method for identifying
CC   modulators of quorum sensing signalling in Pseudomonas aeruginosa
CC   bacteria. Modulators of quorum signaling may be used to treat P.
CC   aeruginosa infections. P. aeruginosa is an opportunistic pathogen of
CC   immunocompromised individuals (burn patients, cystic fibrosis patients,
CC   patients undergoing immunosuppressive therapy and patients with AIDS).
XX
SQ   Sequence 1257 BP; 210 A; 469 C; 402 G; 176 T; 0 other;

  Query Match             11.4%;  Score 143.8;  DB 22;  Length 1257;
  Best Local Similarity   48.7%;  Pred. No. 1.5e-23;
  Matches  455;  Conservative    0;  Mismatches  447;  Indels   33;  Gaps    1;

Qy        254 CCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCGGAGGATCACGCTC 313
Db        233 CGGGGATCGGCGAGGGACTCAGCCACCACATGCTCAACCTCGACCCGCCGGACCATACCC 292

Qy        314 GGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGCGCGCCG 373
Db        293 GCCTGCGCTCGCTGGTTGGCCGCGCGTTCACCCCGCGCCAGGTGGAGCGCCTGCAACCGC 352

Qy        374 AAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACG 433
Db        353 ATATAGAACGGATCACCGAGGCATTGCTGGACGCCATGGCCGGCCGCGAACAGGCCGACC 412

Qy        434 TTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTC 493
Db        413 TGATGGCCGACTTCGCGATCCCGCTGACCATCGCGGTGATCTTCGAGCTGCTGGGCATTC 472

Qy        494 CGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCG 553
Db        473 CCGAGGCCGAGCGCGAACACGCCCGCCAGTCCTG------------------------- 506
```

```
Qy        554 TGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGG 613
Db        507 -------GGAGCGCCAGGCGGAACTGCTGTCGCCGGAGGAGGCCCAGGCCCTGGCCGATG 559

Qy        614 GGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACG 673
Db        560 CGCAGGTCGACTACCTGCGCGTGCTGCTCGAGGCCAAGCGCCGGCAGCCGGCCGACGACG 619

Qy        674 TCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGG 733
Db        620 TCTACAGCGGGCTGGTGCAGGCCGCCGACGAGAGCGGCCAGTTGAGCGAAGCGGAACTCG 679

Qy        734 TCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCGT 793
Db        680 TCTCCATGGCCCACCTGCTGATGATGAGCGGCTTCGAGACCACCATGAACATGATCGGCA 739

Qy        794 TCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCG 853
Db        740 ACGCGCTGGTCACCCTGCTGGTCAACCCGGAGCAACTGGCGTTGCTGCGGGCGCAGCCGG 799

Qy        854 GGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAATAGGAA 913
Db        800 AACTCCTGCCCAACGCCATGGAAGAACTGGTCCGCCACGACAGTCCGGTGCGCGCCTCGA 859

Qy        914 CTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGGAGA 973
Db        860 TGTTGCGCTTCACCGTGGAAGACGTGGAACTGGACGGGGTCACCATTCCCGCCGGCGAAT 919

Qy        974 TGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGACTGTATTCTCCAGGCCAGACG 1033
Db        920 ACATCCTGGTCTCCAACCTGACCGCCAACCACGATGCCGAGCGCTTCGACGATCCCGACC 979

Qy       1034 TGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCCCCATGTCT 1093
Db        980 GCCTCGACCTCACCCGCAACACCGATGGCCATCTCGGCTACGGCTTCGGCGTGCACTACT 1039

Qy       1094 GCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCACCATCTTCCGTA 1153
Db       1040 GCGTCGGCGCCTCGCTGGCCCGGCTGGAGGGGCGGATCGCCATCCAGCGCCTGCTCGCGC 1099

Qy       1154 GGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTG 1188
Db       1100 GCTTCCCCGACCTCCAGTTGGCGGTGCCCCACGCG 1134
```

RESULT 5
AAQ14548
ID   AAQ14548 standard; DNA; 1400 BP.
XX
AC   AAQ14548;
XX
DT   28-JAN-1992 (first entry)
XX
DE   EryF gene.
XX
KW   C-6 hydroxylation; erythromycin; 6-deoxyerythromycin; antibiotics;
KW   saccharopolyspora; cytochrome P450 monooxygenase; ss.
XX
OS   Saccaropolyspora erythraea.
XX
FH   Key   Location/Qualifiers
FT   CDS   158..1372
FT   /*tag= a
XX
PN   WO9116334-A.
XX
PD   31-OCT-1991.
XX
PF   16-APR-1991;   91WO-US02600.
XX
PR   18-APR-1990;   90US-0510483.
XX
PA   (ABBO) ABBOTT LABORATORIES.
XX
PI   Weber JM;
XX
DR   WPI; 1991-339744/46.
DR   P-PSDB; AAR14724.
XX
PT   New 6-deoxy:erythromycin derivs. - are antibiotics with increased
PT   acid stability, produced by cultivation of saccharopolyspora.
XX
PS   Disclosure; Fig 3; 56pp; English.
XX
CC   The eryF gene encodes the 6-hydroxylase component of the cytochrome
CC   P450 monooxygenase system responsible for the hydroxylation of 6-de-
CC   oxyerythronolide B to erythronolide B. Interruption of this step
CC   results in the formation of deoxyerythromycin A and new derivatives
CC   useful as antibiotic which have better stability against acids that
CC   the corresponding erythromycins. Interruption of the reaction can
CC   be effected by an insertion into the eryF gene of a plasmid, gene
CC   replacement or chemical or light-induced mutagenesis.
CC   The gene is positioned between the eryH and eryG genes.
CC   See also AAQ14549.
XX
SQ   Sequence 1400 BP; 212 A; 495 C; 479 G; 214 T; 0 other;

```
  Query Match             10.4%;  Score 131.6;  DB 12;  Length 1400;
  Best Local Similarity   49.1%;  Pred. No. 9.5e-21;
  Matches  430;  Conservative    0;  Mismatches  419;  Indels   27;  Gaps    2;

Qy        298 CCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATC 357
Db        440 CCGCCGACCCACACCCGGCTGCGCAAGCTGGTGTCGCAGGAGTTCACCGTCCGCCGCGTG 499

Qy        358 GACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGA 417
Db        500 GAGGCGATGCGGCCCCGCGTCGAGCAGATCACCGCGGAGCTGCTCGACGAGGTGGGCGAC 559

Qy        418 CAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGC 477
Db        560 TCCGGCGTGGTCGACATCGTCTACCGCTTCGCCCACCCGCTGCCCATCAAGGTCATCTGC 619

Qy        478 GCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACT 537
Db        620 GAGCTGCTCGGCGTCGACGAGAAGTACCGCGGGGAGTTCGGCGGTGGAGCTCGGAGA-- 677

Qy        538 GCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTC 597
Db        678 ---------------------TCCTGGTCATGGACCCGGAGCGGGCCGAACAGCGCGGG 715

Qy        598 GCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAAC 657
Db        716 CAGGCGGCCAGGGAGGTCGTCAACTTCATCCTCGACCTGGTCGAGCGCCGCCGCACCGAG 775

Qy        658 CCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTG 717
Db        776 CCCGGCGACGACCTGCTGTCCGCGCTTATCAGGGTCCAGGACGACGATGACGGTCGGCTC 835

Qy        718 AGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACG 777
Db        836 AGCGCCGACGAGCTGTCCTCCATCGCGCTGGTGCTGCTGCTGGCCGGTTTCGAGGCGTCG 895

Qy        778 ATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTG 837
Db        896 GTGAGCCTCATCGGGATCGGCACCTACCTGCTGCTCACCCACCCGGACCAGCTCGCGCTG 955

Qy        838 GTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAAT 897
Db        956 GTGCGGCGGGACCCGTCGGCGCTGCCCAACGCCGTCGAGGAGATCCTGCGCTACATCGCT 1015

Qy        898 ATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCG 957
Db       1016 CTGCAGGAGAC---CACCACGCGCTTCGCCGCGGAGGACCTGGAGATCGGCGGTGTCGCG 1072
```

```
Qy        958 ATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTA 1017
Db       1073 ATCCCCAGTACAGCACGGTGCTGGTCGCGAACGGCGCGGCCAACCGCGACCCGAAGCAG 1132

Qy       1018 TTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGT 1077
Db       1133 TTCCCGGACCCCCACCGCTCCGACGTCACCCGCGACACCCGCGGCCACCTGTCGTTCGGG 1192

Qy       1078 AGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTG 1137
Db       1193 CAGGGCATCCACTTCTGCATGGGCCGGCCGCTGGCCAAGCTGGAGGGCGAGGTGGCGCTG 1252

Qy       1138 GGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTG 1173
Db       1253 CGGGCGCTGTTCGGCCGCTTCCCCGCTCTGTCGCTG 1288
```

RESULT 6
ABX56047
ID   ABX56047 standard; DNA; 1347 BP.
XX
AC   ABX56047;
XX
DT   13-FEB-2003 (first entry)
XX
DE   M. echinospora calicheamicin biosynthesis gene calW.
XX
KW   Calicheamicin biosynthetic gene cluster; aryltetrasaccharide;
KW   aglycone; calicheamicin resistance; nonchromoprotein enediyne;
KW   enediyne resistance; bone marrow cell; gene therapy; gene; ds.
XX
OS   Micromonospora echinospora spp. calichensis.
XX
PN   WO200279465-A2.
XX
PD   10-OCT-2002.
XX
PF   28-NOV-2001; 2001WO-US44285.
XX
PR   28-NOV-2000; 2000US-0724797.
XX
PA   (SLOK) SLOAN KETTERING INST CANCER RES.
XX
PI   Thorson J;
XX
DR   WPI; 2003-092897/08.
DR   P-PSDB; ABU11985.
XX
PT   Novel nucleic acid molecule from nonchromoprotein enediyne biosynthetic
PT   gene cluster from Micromonospora echinospora useful for conferring
PT   calicheamicin resistance on a subject -
XX
PS   Claim 9; Page 120-122; 179pp; English.
XX
CC   The present invention relates to the isolation of the Micromonospora
CC   echinospora spp. calichensis calicheamicin biosynthetic gene cluster
CC   encoding proteins and enzymes used in the biosynthetic production of
CC   calicheamicin, including aryltetrasaccharide and aglycone. The gene
CC   cluster also includes the gene encoding for the protein that confers
CC   calicheamicin resistance. The calicheamicin biosynthetic gene cluster
CC   is a nonchromoprotein enediyne biosynthetic gene cluster. Expression
CC   vectors containing genes from the biosynthetic gene cluster are also
CC   disclosed. The expression vectors are useful for producing calicheamicin
CC   biosynthetic proteins. The calicheamicin self-resistance gene provides
CC   an approach for gene therapy, for example, by introduction of enediyne
CC   resistance genes into bone marrow cells, thus increasing resistance and
CC   allowing tolerance to chemotherapeutic doses of calicheamicin.
CC   ABX56026-ABX56073 represent genes from the M. echinospora calicheamicin
CC   biosynthesis gene cluster.
XX
SQ   Sequence 1347 BP; 163 A; 505 C; 513 G; 166 T; 0 other;

Query Match 10.2%; Score 128.8; DB 25; Length 1347;
Best Local Similarity 49.7%; Pred. No. 4.2e-20;
Matches 356; Conservative 0; Mismatches 357; Indels 3; Gaps 1;

```
Qy        459 CCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCG 518
Db        558 CCTGGTGCCGTCGTTCAGCTATCCGGTGCCGTCGCGGGTGATCTGCGAGATGCTCGGCGT 617

Qy        519 TCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGA 578
Db        618 GCCGTACGGCGAACACGCGTTCTTCGAGCGCCGGTCCACCCGGATGCTGAGTCGCGGCGT 677

Qy        579 GGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCT 638
Db        678 GCCCGCGGACGAGCGGGCCCGGTGCGCCCGCGAGATCCGCGAGTTCCTCGACGGCGTGGT 737

Qy        639 CGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGA 698
Db        738 GACCGACAAGGAGCGGCACCC---CGGCGACGACGTGCTCAGCCGGCTGCTCGCCGCGCA 794

Qy        699 GGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGC 758
Db        795 GCGCGCGGCCGGCGAGCCCGACCACGAGGCCGTGGTGAGCATGGCCTTCGTGCTGCTGGT 854

Qy        759 TGCTGGCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTC 818
Db        855 CGCCGGGCACGTCACGACGTCGAACATGATCTCGCTGAGCGTGCTGGCCCTGTTGACCCA 914

Qy        819 GCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGA 878
Db        915 TCCGGAGCGGCTCGCCCGCCTGCGCGCCGAGCCGGACCGGTTCCCCGCCGCCGTCGAGGA 974

Qy        879 GGTGCTCCGCTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCT 938
Db        975 GCTGCTGCGGTACTTCACCATCGTCGAGGCGGCGACCGCGCGGACCGCGACCGCCGACGT 1034

Qy        939 GGAGTACTGCGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGC 998
Db       1035 GACGGTCGGTGGGGTCACCATCCGGGCCGGGGAGGGGGTGGTGGCGCTGGGCCAGGCCGC 1094

Qy        999 CCTGAGAGATGGGACTGTATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGG 1058
Db       1095 CAACCGGGACCCGGCGGCGTTCGACCGGCCGGACGAGTTCGACCCGGACCGCGACGCCCG 1154

Qy       1059 CGCGAGCCTCGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCT 1118
Db       1155 GCACCACCTCGCCTTCGGCTACGGACGACACATCTGCCCCGGCCAGCACCTGGCCCGGCT 1214

Qy       1119 CGAGGCGGAGATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGA 1174
Db       1215 GGAACTGGACGTCGCGCTGAGCCGGCTGGTCCGGCGGCTGCCCGGGCTGCGGTTGA 1270
```

RESULT 7
AAA75635
ID   AAA75635 standard; DNA; 5970 BP.
XX
AC   AAA75635;
XX
DT   22-JAN-2001 (first entry)
XX
DE   Nucleotide sequence of ORF12 which encodes a transcriptional activator.
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis; ss.
XX
OS   Streptomyces venezuelae.
XX
PN   US6117659-A.
XX
PD   12-SEP-2000.
XX

```
PF   27-MAY-1999;   99US-0320878.
XX
PR   28-MAY-1998;   98US-0087080.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
PR   20-MAY-1999;   99US-0134990.
PR   30-APR-1997;   97US-0846247.
PR   06-MAY-1998;   98US-0073538.
PR   28-AUG-1998;   98US-0141908.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  Tang L,  McDaniel R;
XX
DR   WPI; 2000-610844/58.
XX
PT   New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT   for converting ketolides to antibiotics and as antibiotics and
PT   intermediates in the synthesis of compounds with pharmaceutical value
PT   -
XX
PS   Disclosure; Columns 41-44; 117pp; English.
XX
CC   The present sequence is used to produce the recombinant DNA compounds
CC   of the invention. The specification describes a recombinant DNA compound
CC   expressing recombinant polyketide synthase genes in host cells for the
CC   production of narbonolide, narbonolide derivatives and polyketides that
CC   are useful as antibiotics and as intermediates in the synthesis of
CC   compounds with pharmaceutical value. The DNA compounds may also encode
CC   a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC   transferase enzymes (useful for conversion of ketolides to antibiotics),
CC   and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC   These compounds are also useful for increasing the antibiotic activity
CC   of a compound relative to the unhydroxylated compound. The recombinant
CC   host cells are useful as genetic systems that allow rapid engineering
CC   of the narbonolide polyketide synthase. These would be valuable for
CC   creating novel ketolide analogs for pharmaceutical applications.
XX
SQ   Sequence 5970 BP; 806 A; 2219 C; 2142 G; 800 T; 3 other;

  Query Match             9.8%;  Score 123;  DB 21;  Length 5970;
  Best Local Similarity   47.5%;  Pred. No. 1.2e-18;
  Matches  452;  Conservative    0;  Mismatches  475;  Indels   24;  Gaps    2;

Qy        236 AGTACTCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGC 295

Db       1582 ACTCCACGACTCCCCTGACCGAGGCCGAGGCCGCGCTCAACCACAACATGCTGGAGTCCG 1641

Qy        296 CGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCA 355

Db       1642 ACCCGCCGCGGCACACCCGGCTGCGCAAGCTGGTGGCCCGTGAGTTCACCATGCGCCGGG 1701

Qy        356 TCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCG 415

Db       1702 TCGAGTTGCTGCGGCCCCGGGTCCAGGAGATCGTCGACGGGCTCGTGGACGCCATGCTGG 1761

Qy        416 GACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCA 475

Db       1762 CGGCGCCCGACGGCCGCGCCGATCTGATGGAGTCCCTGGCCTGGCCGCTGC---CGATCA 1818

Qy        476 GCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGA 535

Db       1819 CCGTGATCTCCGAACTCCTCGGCGTGCCCGAGCCGGACCGCGCCGCCTTCCGCGTCTGGA 1878

Qy        536 CTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGG 595

Db       1879 CCGACGCCTTCGTCTTCCCGGACGATCCCGCCCAGGCCCAGACCG--------------- 1923

Qy        596 TCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGA 655

Db       1924 ------CCATGGCCGAGATGAGCGGCTATCTCTCCCGGCTCATCGACTCCAAGCGCGGGC 1977

Qy        656 ACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGC 715

Db       1978 AGGACGGCGAGGACCTGCTCAGCGCGCTCGTGCGGACCAGCGACGAGGACGGCTCCCGGC 2037

Qy        716 TGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCA 775

Db       2038 TGACCTCCGAGGAGCTGCTCGGTATGGCCCACATCCTGCTCGTCGCGGGGCACGAGACCA 2097

Qy        776 CGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGC 835

Db       2098 CGGTCAATCTGATCGCCAACGGCATGTACGCGCTGCTCTCGCACCCCGACCAGCTGGCCG 2157

Qy        836 TGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACA 895

Db       2158 CCCTGCGGGCCGACATGACGCTCTTGGACGGCGCGGTGGAGGAGATGTTGCGCTACGAGG 2217

Qy        896 ATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCAT 955

Db       2218 GCCCGGTGGAATCCGCGACCTACCGCTTCCCGGTCGAGCCCGTCGACCTGGACGGCACGG 2277

Qy        956 CGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGACTG 1015

Db       2278 TCATCCCGGCCGGTGACACGGTCCTCGTCGTCCTGGCCGACGCCCACCGCACCCCCGAGC 2337

Qy       1016 TATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACG 1075

Db       2338 GCTTCCCGGACCCGCACCGCTTCGACATCCGCCGGGACACCGCCGGCCATCTCGCCTTCG 2397

Qy       1076 GTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCG 1135

Db       2398 GCCACGGCATCCACTTCTGCATCGGCGCCCCCTTGGCCCGGTTGGAGGCCCGGATCGCCG 2457

Qy       1136 TGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCG 1186

Db       2458 TCCGCGCCCTTCTCGAACGCTGCCCGGACCTCGCCCTGGACGTCTCCCCCG 2508


RESULT 8
AAZ56003
ID   AAZ56003 standard; DNA; 5970 BP.
XX
AC   AAZ56003;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Contig 002 from cosmid pKOS023-27 from Streptomyces venezuelae.
XX
KW   Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; contig 002;
KW   ketolide; amino transferase dehydrase; hydroxylase; picromycin;
KW   antibiotic production; narbomycin; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..995)
FT                   /*tag=  a
FT                   /product= PICCIV
FT                   /note= "Partial amino transferase-dehydrase"
FT                   /transl_except= (Pos:180..182, aa:Xaa)
FT   CDS             1356..2606
FT                   /*tag=  c
FT                   /product= PicK
FT                   /note= "Cytochrome P450 hydroxylase"
FT   CDS             2739..5525
FT                   /*tag=  d
FT                   /product= transcriptional_activator
FT                   /transl_except= (Pos:4818..4820, aa:Xaa)
FT                   /note= "Xaa = unknown"
XX
PN   WO9961599-A2.
XX
PD   02-DEC-1999.
XX
PF   27-MAY-1999;   99WO-US11814.
```

```
XX
PR   28-MAY-1998;   98US-0087080.
PR   28-AUG-1998;   98US-0141908.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-072618/06.
DR   P-PSDB; AAY67209, AAY67213, AAY67218.
XX
PT   New recombinant DNA encoding a domain of narbonolide polyketide
PT   synthase, for production of ketolide antibiotics -
XX
PS   Disclosure; Page 35-37; 98pp; English.
XX
CC   This is contig 002 from the recombinant cosmid pKOS023-27 DNA sequence
CC   (see AAZ56001) which contains a Streptomyces venezuelae DNA insert. The
CC   cosmid contains open reading frames which encode the various modules of
CC   the narbonolide polyketide synthase (PKS). The invention relates to
CC   recombinant DNA containing a coding sequence for a narbonolide PKS.
CC   Polyketides are compounds synthesised from 2-carbon units through a
CC   series of condensations and subsequent modifications. Modular PKSs are
CC   responsible for the production of many antibiotics including picromycin.
CC   The narbonolide PKS consists of a loading module, six extender modules,
CC   and two thioester domains. Four proteins make up the narbonolide PKS
CC   (PICAI, PICAII, PICAIII and PICAIV). PICAI includes the loading module
CC   and extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC   PICAIII includes extender module 5 and PICAIV includes extender module 6
CC   and a type II thioesterase domain. The second type II thioesterase
CC   domain is found on the PICB protein. The nucleotide sequences encoding
CC   all of these proteins can be isolated in recombinant form from the
CC   recombinant cosmid pKOS023-27. Narbonolide is desosaminylated in S.
CC   venezuelae to yield narbomycin, the desosaminyl transferase enzyme is
CC   required for this conversion, and the desosamine biosynthetic genes are
CC   also found in cosmid pKOS023-27. The recombinant DNA of the invention is
CC   used to express, in transformed cells, narbonolide (or its derivatives)
CC   or other ketolides (particularly hybrids), which may then be converted
CC   (e.g. by other enzymes recombinantly expressed in the same hosts) to
CC   polyketide antibiotics or their intermediates. The antibiotics are useful
CC   in human or veterinary medicine.
XX
SQ   Sequence 5970 BP; 806 A; 2219 C; 2142 G; 800 T; 3 other;

  Query Match             9.8%;  Score 123;  DB 21;  Length 5970;
  Best Local Similarity   47.5%;  Pred. No. 1.2e-18;
  Matches  452;  Conservative    0;  Mismatches  475;  Indels   24;  Gaps    2;

Qy        236 AGTACTCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGC 295
Db       1582 ACTCCACGACTCCCCTGACCGAGGCCGAGGCCGCGCTCAACCACAACATGCTGGAGTCCG 1641

Qy        296 CGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCA 355
Db       1642 ACCCGCCGCGGCACACCCGGCTGCGCAAGCTGGTGGCCCGTGAGTTCACCATGCGCCGGG 1701

Qy        356 TCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCG 415
Db       1702 TCGAGTTGCTGCGGCCCCGGGTCCAGGAGATCGTCGACGGGCTCGTGGACGCCATGCTGG 1761

Qy        416 GACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCA 475
Db       1762 CGGCGCCCGACGGCCGCGCCGATCTGATGGAGTCCCTGGCCTGGCCGCTGC---CGATCA 1818

Qy        476 GCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGA 535
Db       1819 CCGTGATCTCCGAACTCCTCGGCGTGCCCGAGCCGGACCGCGCCGCCTTCCGCGTCTGGA 1878

Qy        536 CTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGG 595
Db       1879 CCGACGCCTTCGTCTTCCCGGACGATCCCGCCCAGGCCCAGACCG--------------- 1923

Qy        596 TCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGA 655
Db       1924 ------CCATGGCCGAGATGAGCGGCTATCTCTCCCGGCTCATCGACTCCAAGCGCGGGC 1977

Qy        656 ACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGC 715
Db       1978 AGGACGGCGAGGACCTGCTCAGCGCGCTCGTGCGGACCAGCGACGAGGACGGCTCCCGGC 2037

Qy        716 TGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCA 775
Db       2038 TGACCTCCGAGGAGCTGCTCGGTATGGCCCACATCCTGCTCGTCGCGGGGCACGAGACCA 2097

Qy        776 CGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGC 835
Db       2098 CGGTCAATCTGATCGCCAACGGCATGTACGCGCTGCTCTCGCACCCCGACCAGCTGGCCG 2157

Qy        836 TGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACA 895
Db       2158 CCCTGCGGGCCGACATGACGCTCTTGGACGGCGCGGTGGAGGAGATGTTGCGCTACGAGG 2217

Qy        896 ATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCAT 955
Db       2218 GCCCGGTGGAATCCGCGACCTACCGCTTCCCGGTCGAGCCCGTCGACCTGGACGGCACGG 2277

Qy        956 CGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTG 1015
Db       2278 TCATCCCGGCCGGTGACACGGTCCTCGTCGTCCTGGCCGACGCCCACCGCACCCCCGAGC 2337

Qy       1016 TATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACG 1075
Db       2338 GCTTCCCGGACCCGCACCGCTTCGACATCCGCCGGGACACCGCCGGCCATCTCGCCTTCG 2397

Qy       1076 GTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCG 1135
Db       2398 GCCACGGCATCCACTTCTGCATCGGCGCCCCCTTGGCCCGGTTGGAGGCCCGGATCGCCG 2457

Qy       1136 TGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCG 1186
Db       2458 TCCGCGCCCTTCTCGAACGCTGCCCGGACCTCGCCCTGGACGTCTCCCCCG 2508


RESULT 9
ABS56092
ID   ABS56092 standard; DNA; 5970 BP.
XX
AC   ABS56092;
XX
DT   21-JAN-2003  (first entry)
XX
DE   Contig 002 from cosmid pKOS023-27S containing S. venezuelae DNA insert.
XX
KW   Narbonolide polyketide synthase; PKS; desosamine biosynthetic gene;
KW   desosaminyl transferase gene; beta-glucosidase gene; antibiotic;
KW   picK hydroxylase gene; C12 hydroxylase gene; narbonolide;
KW   desosaminylated polyketide; narbomycin biosynthesis; mutant;
KW   picromycin biosynthesis; ds.
XX
OS   Streptomyces venezuelae.
OS   Synthetic.
XX
PN   WO200297062-A2.
XX
PD   05-DEC-2002.
XX
PF   22-FEB-2002; 2002WO-US05642.
XX
PR   22-FEB-2001; 2001US-0793708.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  McDaniel R,  Tang L;
XX
```

```
DR   WPI; 2003-041412/03.
XX
PT   Preparation of polyketides by recombinant DNA technology, useful as
PT   antibiotics and as intermediates in the synthesis of pharmaceutical
PT   compounds -
XX
PS   Disclosure; Page 40-41; 127pp; English.
XX
CC   The present invention relates to recombinant DNA sequences encoding
CC   for a narbonolide polyketide synthase (PKS) domain, and methods of
CC   producing polyketides by recombinant DNA technology. The recombinant
CC   DNA sequences are derived from Streptomyces venezuelae desosamine
CC   biosynthetic, desosaminyl transferase, beta-glucosidase, or picK
CC   (C12) hydroxylase genes. The method is useful for transforming a cell
CC   with a recombinant expression vector that encodes a functional
CC   beta-glucosidase gene, and therefore for increasing the yield of a
CC   desosaminylated polyketide in a cell. The recombinant methods and
CC   materials are useful for expressing polyketides with significant
CC   antibiotic activity, derived in whole or in part from the narbonolide
CC   PKS gene, and other genes involved in narbomycin and picromycin
CC   biosynthesis in recombinant host cells. The present sequence
CC   represents contig 002 from cosmid pKOS023-27S containing
CC   S. venezuelae DNA.
XX
SQ   Sequence 5970 BP; 806 A; 2219 C; 2142 G; 800 T; 3 other;

  Query Match             9.8%;  Score 123;  DB 25;  Length 5970;
  Best Local Similarity   47.5%;  Pred. No. 1.2e-18;
  Matches  452;  Conservative    0;  Mismatches  475;  Indels   24;  Gaps    2;

Qy        236 AGTACTCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGC 295
Db       1582 ACTCCACGACTCCCCTGACCGAGGCCGAGGCCGCGCTCAACCACAACATGCTGGAGTCCG 1641

Qy        296 CGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCA 355
Db       1642 ACCCGCCGCGGCACACCCGGCTGCGCAAGCTGGTGGCCCGTGAGTTCACCATGCGCCGGG 1701

Qy        356 TCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCG 415
Db       1702 TCGAGTTGCTGCGGCCCCGGGTCCAGGAGATCGTCGACGGGCTCGTGGACGCCATGCTGG 1761

Qy        416 GACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCA 475
Db       1762 CGGCGCCCGACGGCCGCGCCGATCTGATGGAGTCCCTGGCCTGGCCGCTGC---CGATCA 1818

Qy        476 GCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGA 535
Db       1819 CCGTGATCTCCGAACTCCTCGGCGTGCCCGAGCCGGACCGCGCCGCCTTCCGCGTCTGGA 1878

Qy        536 CTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGG 595
Db       1879 CCGACGCCTTCGTCTTCCCGGACGATCCCGCCCAGGCCCAGACCG--------------- 1923

Qy        596 TCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGA 655
Db       1924 ------CCATGGCCGAGATGAGCGGCTATCTCTCCCGGCTCATCGACTCCAAGCGCGGGC 1977

Qy        656 ACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGC 715
Db       1978 AGGACGGCGAGGACCTGCTCAGCGCGCTCGTGCGGACCAGCGACGAGGACGGCTCCCGGC 2037

Qy        716 TGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCA 775
Db       2038 TGACCTCCGAGGAGCTGCTCGGTATGGCCCACATCCTGCTCGTCGCGGGGCACGAGACCA 2097

Qy        776 CGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGC 835
Db       2098 CGGTCAATCTGATCGCCAACGGCATGTACGCGCTGCTCTCGCACCCCGACCAGCTGGCCG 2157

Qy        836 TGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACA 895
Db       2158 CCCTGCGGGCCGACATGACGCTCTTGGACGGCGCGGTGGAGGAGATGTTGCGCTACGAGG 2217

Qy        896 ATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCAT 955
Db       2218 GCCCGGTGGAATCCGCGACCTACCGCTTCCCGGTCGAGCCCGTCGACCTGGACGGCACGG 2277

Qy        956 CGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTG 1015
Db       2278 TCATCCCGGCCGGTGACACGGTCCTCGTCGTCCTGGCCGACGCCCACCGCACCCCCGAGC 2337

Qy       1016 TATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACG 1075
Db       2338 GCTTCCCGGACCCGCACCGCTTCGACATCCGCCGGGACACCGCCGGCCATCTCGCCTTCG 2397

Qy       1076 GTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCG 1135
Db       2398 GCCACGGCATCCACTTCTGCATCGGCGCCCCCTTGGCCCGGTTGGAGGCCCGGATCGCCG 2457

Qy       1136 TGGGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCG 1186
Db       2458 TCCGCGCCCTTCTCGAACGCTGCCCGGACCTCGCCCTGGACGTCTCCCCCG 2508

RESULT 10
ABN88910
ID   ABN88910 standard; DNA; 1191 BP.
XX
AC   ABN88910;
XX
DT   21-AUG-2002  (first entry)
XX
DE   Streptomyces sp. cytochrome P450 encoding DNA SEQ ID NO:1.
XX
KW   Streptomyces; cytochrome P450; anticancer; UCN-01; monooxygenase;
KW   microorganism; gene; ds.
XX
OS   Streptomyces sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1191
FT                   /*tag=  a
FT                   /product= "cytochrome P450"
XX
PN   JP2002058490-A.
XX
PD   26-FEB-2002.
XX
PF   22-AUG-2000; 2000JP-0251304.
XX
PR   22-AUG-2000; 2000JP-0251304.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
DR   WPI; 2002-410015/44.
DR   P-PSDB; ABB81315.
XX
PT   Novel cytochrome P450 used for improving the productivity of anticancer
PT   drug UCN-01  -
XX
PS   Claim 5; Page 9-11; 22pp; Japanese.
XX
CC   The present invention describes 3 Streptomyces cytochrome P450 (I)
CC   sequences, ABB81315 to ABB81317 encoded by ABN88910 to ABN88912
CC   respectively. (I) have monooxygenase activity and improve the
CC   productivity of an anticancer drug designated UCN-01. (I) is used
CC   for the production of UCN-01 with transformed Streptomyces sp.
CC   microorganisms.
XX
SQ   Sequence 1191 BP; 169 A; 423 C; 431 G; 168 T; 0 other;

  Query Match             9.7%;  Score 122.6;  DB 24;  Length 1191;
  Best Local Similarity   48.3%;  Pred. No. 1.1e-18;
  Matches  430;  Conservative    0;  Mismatches  434;  Indels   27;  Gaps    2;
```

```
Qy        298 CCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATC 357
Db        271 CCGCCCCAGCACACCCGGCTGCGCGGCTGATCGCCCGGGAGTTCACCGGACGCCGGGTC 330

Qy        358 GACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGA 417
Db        331 GAGCGGCTGCGGCCGAGGGTCCAGGAGATCACCGACTCGCTGCTGGACGAGATGCTGCCG 390

Qy        418 CAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGC 477
Db        391 CGCGGCCGCGCCGACCTGGTCGAGTCCTTCGCGTACCCGCTGCCCCTCACCGTCATCTGC 450

Qy        478 GCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACT 537
Db        451 GAACTCCTCGGGGTGCCCGAGA-----------------------TCGACCGGGCGGCC 486

Qy        538 GCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTC 597
Db        487 TTCCGCAAGCTGTCGACGGAGGCGGTGGCACCCACCAGCGGGGAGAGCGAGTAC---GCG 543

Qy        598 GCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAAC 657
Db        544 GCCTTCGTCCAACTCGCCGCCTACCTGGAGGAGTTGGTCGAGGAGAAGCGGTGCGCACCG 603

Qy        658 CCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTG 717
Db        604 CCGGCCGACGATCTGCTGAGCGCGCTGATCCGGACGACCGACGAGGACGGCGACCGTCTG 663

Qy        718 AGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACG 777
Db        664 TCACCCGCGGAGCTGCGCGGCATGGCCTTCATCCTGCTGATCGCCGGCCACGAGACCACC 723

Qy        778 ATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTG 837
Db        724 GTCAACCTCATCACCGGTGCCGTCCACGCCCTCCTCACGCATCCCGGGCAACTCGCCCAG 783

Qy        838 GTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAAT 897
Db        784 GTGCGGGGCGACATGAGCCTCGTGGACGCGGTCGTGGAGGAGACGCTGCGCCATGAGGGG 843

Qy        898 ATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCG 957
Db        844 CCGGTGGAGAACGCGACGTTCCGCTTCGCCGCCGAGCCGCTGGAGATAGGGGGCACGGTC 903

Qy        958 ATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTA 1017
Db        904 ATCCCAGCGGGGGACCCGGTGCTGATCGGGCTGGCCGCGGCCGATCGGGACGGCGCCCGC 963

Qy       1018 TTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGT 1077
Db        964 TATCCCGGCCCCGACCGCTTCGACATCCACCGGGACACACGCGGGCATCTCGCCTTCGGC 1023

Qy       1078 AGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTG 1137
Db       1024 CACGGCATCCACTTCTGCCTCGGCGCGCCGCTCGCCCGGCTGGAGGCCCGGGTGGCTCTG 1083

Qy       1138 GGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTG 1188
Db       1084 CGCGCGCTGCTGGAGCGGTGCCCTGGTCTCACGCCGGACGGGGCACCGGGG 1134
```

```
RESULT 11
AAZ87301
ID   AAZ87301 standard; DNA; 1251 BP.
XX
AC   AAZ87301;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide biosynthetic gene pikC, SEQ ID NO:38.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1251
FT                   /*tag=  a
FT                   /product= "PikC"
FT                   /transl_except= (pos:307..309, aa:Leu)
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77196.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 15; Page 428-429; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAZ87295-Z87302 represent macrolide biosynthetic genes from
CC   Streptomyces venezuelae ATCC 15439, which encode proteins
CC   AAY77190-Y77197.
XX
SQ   Sequence 1251 BP; 171 A; 484 C; 413 G; 183 T; 0 other;

  Query Match              9.6%;  Score 121.4;  DB 21;  Length 1251;
  Best Local Similarity   50.4%;  Pred. No. 2.1e-18;
  Matches  295;  Conservative    0;  Mismatches  290;  Indels    0;  Gaps    0;

Qy        602 CCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGC 661
Db        569 CCATGGCCGAGATGAGCGGCTATCTCTCCCGGCTCATCGACTCCAAGCGCGGGCAGGACG 628

Qy        662 TCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCA 721
```

```
Db        629 GCGAGGACCTGCTCAGCGCGCTCGTGCGGACCAGCGACGAGGACGGCTCCCGGCTGACCT 688

Qy        722 CGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCT 781

Db        689 CCGAGGAGCTGCTCGGTATGGCCCACATCCTGCTCGTCGCGGGGCACGAGACCACGGTCA 748

Qy        782 ACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGA 841

Db        749 ATCTGATCGCCAACGGCATGTACGCGCTGCTCTCGCACCCCGACCAGCTGGCCGCCCTGC 808

Qy        842 AGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCC 901

Db        809 GGGCCGACATGACGCTCTTGGACGGCGCGGTGGAGGAGATGTTGCGCTACGAGGGCCCGG 868

Qy        902 TCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCA 961

Db        869 TGGAATCCGCGACCTACCGCTTCCCGGTCGAGCCCGTCGACCTGGACGGCACGGTCATCC 928

Qy        962 AGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTCT 1021

Db        929 CGGCCGGTGACACGGTCCTCGTCGTCCTGGCCGACGCCCACCGCACCCCCGAGCGCTTCC 988

Qy       1022 CCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAG 1081

Db        989 CGGACCCGCACCGCTTCGACATCCGCCGGGACACCGCCGGCCATCTCGCCTTCGGCCACG 1048

Qy       1082 GCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCA 1141

Db       1049 GCATCCACTTCTGCATCGGCGCCCCCTTGGCCCGGTTGGAGGCCCGGATCGCCGTCCGCG 1108

Qy       1142 CCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCG 1186

Db       1109 CCCTTCTCGAACGCTGCCCGGACCTCGCCCTGGACGTCTCCCCCG 1153
```

RESULT 12

```
ABZ76672
ID   ABZ76672 standard; DNA; 1992 BP.
XX
AC   ABZ76672;
XX
DT   30-APR-2003  (first entry)
XX
DE   Streptomyces sp. TM-7 related compactin DNA SEQ ID NO:1.
XX
KW   Streptomyces sp. TM-7; pravastatin; compactin; hyperlipidaemia;
KW   antilipaemic; microorganism; gene; ds.
XX
OS   Streptomyces sp. TM-7.
XX
FH   Key             Location/Qualifiers
FT   CDS             544..1761
FT                   /*tag=  a
FT                   /product= "compactin related protein (ABP95987)"
FT   CDS             1782..1973
FT                   /*tag=  b
FT                   /product= "compactin related protein (ABP95988)"
XX
PN   WO200299109-A1.
XX
PD   12-DEC-2002.
XX
PF   30-MAY-2002; 2002WO-JP05252.
XX
PR   01-JUN-2001; 2001JP-0166412.
XX
PA   (SAOC ) MERCIAN CORP.
XX
PI   Fujii T,  Hirosue S,  Aritoku Y,  Morimiya T,  Johdo O,  Isshiki K;
XX
DR   WPI; 2003-148672/14.
DR   P-PSDB; ABP95987, ABP95988.
XX
PT   Novel Streptomyces sp. produced polypeptide for hydroxylation of
PT   compactin at 6beta-position and its encoded DNA, applicable in
PT   constructing transformant microbes to synthesize pravastatin for
PT   treating hyperlipedemia  -
XX
PS   Claim 1; Page 45-48; 67pp; Japanese.
XX
CC   The present sequence represents a DNA sequence which contains a base
CC   sequence from bases 544-1758 in the sequence of (I) with 1992 base pairs,
CC   or a DNA hybridisable with the DNA under stringent conditions and
CC   encoding a polypeptide with hydroxylase activity on compactin at
CC   6beta-position. Also described: (1) DNA containing base sequences from
CC   bases 544-1758 and from bases 1782-1970 in the sequence of (I) or a DNA
CC   hybridisable with the DNA under stringent conditions and encoding a
CC   polypeptide with hydroxylase activity on compactin at the 6beta-position;
CC   (2) a polypeptide encoded by any of the DNA or containing an amino acid
CC   sequence based on the polypeptide but with some amino acids deleted,
CC   substituted or added and having hydroxylase activity on compactin at
CC   the 6beta-position; (3) a recombinant DNA obtained by integrating with
CC   any of the DNA; (4) a microorganism transferred with the recombinant DNA;
CC   (5) a process for producing pravastatin by culturing the transformant
CC   microorganism before isolating the culture liquor or cells, and addition
CC   of compactin for reaction to give pravastatin for recovery; and (6)
CC   Streptomyces sp. TM-6 (FERM BP-8002) or TM-7 (FERM BP-8003). (I) has
CC   antilipaemic activity. The polypeptide and its encoded DNA are applicable
CC   in constructing transformant microorganisms to synthesise pravastatin for
CC   treating hyperlipidaemia. With the recombinant microorganisms,
CC   pravastatin can be produced efficiently, with much less 6alpha
CC   hydroxylated epimer formed.
XX
SQ   Sequence 1992 BP; 275 A; 691 C; 726 G; 300 T; 0 other;

  Query Match             9.6%;  Score 120.8;  DB 25;  Length 1992;
  Best Local Similarity   50.7%;  Pred. No. 3.1e-18;
  Matches  290;  Conservative    0;  Mismatches  282;  Indels    0;  Gaps    0;

Qy        602 CCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGC 661

Db       1112 CCGCCCGGGACGACCTGGAGAGCTATCTGGGCGCTCTGGTGGACAGCCTGCGAGGCGAGT 1171

Qy        662 TCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCA 721

Db       1172 CCCGGCCGGGCCTGCTGAGCACGCTCGTCAGGGAGCAGCTGGAGAAGGGCGCGATCGACC 1231

Qy        722 CGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCT 781

Db       1232 GGGAGGAGCTGGTGTCGACGGCGATCCTGCTGCTGGTCGCCGGACACGAGGCGACGGCGT 1291

Qy        782 ACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGA 841

Db       1292 CGATGACGTCGCTCAGCGTCATCACCCTCCTCGAACATCCCGACCAGCACGCCGCGTTGC 1351

Qy        842 AGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCC 901

Db       1352 GCGCCGATCCGTCGCTGGTGCCCGGCGCGGTGGAGGAGCTGCTGCGCTATCTGGCCATCG 1411

Qy        902 TCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCA 961

Db       1412 CCGACATCGCCGGCGGGCGGATCGCGACGGCGGACATCGAGATCGACGGGCAGCGCATCC 1471

Qy        962 AGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTCT 1021

Db       1472 GGGCGGGGGAGGGGGTCATCGTCACCAACTCGATCGCCAACCGCGACGGCTCCGTCTTCG 1531

Qy       1022 CCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAG 1081

Db       1532 CCGACCCGGACGCCTTCGACGTGCGGCGCGAGGCCCGCCACCACCTGGCGTTCGGCTACG 1591

Qy       1082 GCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCA 1141

Db       1592 GGGTGCATCAGTGCCTCGGCCAGAACCTGGCCCGCCTCGAACTGGAGGTCATCCTCACGG 1651

Qy       1142 CCATCTTCCGTAGGTTCCCCGAGATGAAGCTG 1173
```

```
              |  | |||    || | ||||   ||  ||||
Db       1652 CGCTGTTCGAGCGGCTGCCCGGTCTGCGGCTG 1683


RESULT 13
ABV75558/c
ID   ABV75558 standard; DNA; 36538 BP.
XX
AC   ABV75558;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Saccharopolyspora butenyl-spinosyn biosynthetic gene cluster 2.
XX
KW   Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;
KW   metabolite; spinosyn; gene; ds.
XX
OS   Saccharopolyspora sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (114..938)
FT                   /*tag= a
FT                   /product= "busF"
FT                   /note= "No start codon given"
FT   CDS             1389..2561
FT                   /*tag= b
FT                   /product= "busG"
FT   CDS             2601..3353
FT                   /*tag= c
FT                   /product= "busH"
FT   CDS             complement (3359..4546)
FT                   /*tag= d
FT                   /product= "busI"
FT   CDS             4684..6303
FT                   /*tag= e
FT                   /product= "busJ"
FT   CDS             6317..7510
FT                   /*tag= f
FT                   /product= "busK"
FT   CDS             7555..8406
FT                   /*tag= g
FT                   /product= "busL"
FT                   /note= "No start codon given"
FT   CDS             8640..9572
FT                   /*tag= h
FT                   /product= "busM"
FT                   /note= "No start codon given"
FT   CDS             complement (9668..10666)
FT                   /*tag= i
FT                   /product= "busN"
FT   CDS             complement (10675..12135)
FT                   /*tag= j
FT                   /product= "busO"
FT   CDS             complement (12864..14177)
FT                   /*tag= k
FT                   /product= "busP"
FT   CDS             14627..15970
FT                   /*tag= l
FT                   /product= "busQ"
FT                   /note= "No start codon given"
FT   CDS             16008..17144
FT                   /*tag= m
FT                   /product= "busR"
FT                   /note= "No start codon given"
FT   CDS             17168..17917
FT                   /*tag= n
FT                   /product= "busS"
FT   CDS             complement (18520..19932)
FT                   /*tag= o
FT                   /product= "ORF LI"
FT   CDS             complement (19978..20488)
FT                   /*tag= p
FT                   /product= "ORF LII"
FT                   /note= "No start codon given"
FT   CDS             complement (20536..21033)
FT                   /*tag= q
FT                   /product= "ORF LIII"
FT                   /note= "No start codon given"
FT   CDS             21179..21925
FT                   /*tag= r
FT                   /product= "ORF LIV"
FT   CDS             complement (22671..23453)
FT                   /*tag= s
FT                   /product= "ORF LVI"
FT   CDS             complement (23687..24886)
FT                   /*tag= t
FT                   /product= "ORF LVII"
FT   CDS             complement (26177..26923)
FT                   /*tag= u
FT                   /product= "ORF LVIII"
FT                   /note= "No start codon given"
FT   CDS             27646..28476
FT                   /*tag= v
FT                   /product= "ORF LIX"
FT                   /note= "No start codon given"
XX
PN   WO200279477-A2.
XX
PD   10-OCT-2002.
XX
PF   28-MAR-2002; 2002WO-US09968.
XX
PR   30-MAR-2001; 2001US-280175P.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Hahn DR,  Jackson JD,  Bullard BS,  Gustafson GD,  Waldron C;
PI   Mitchell JC;
XX
DR   WPI; 2003-058434/05.
DR   P-PSDB; ABP57683, ABP57684, ABP57685, ABP57686, ABP57687, ABP57688,
DR   ABP57689, ABP57690, ABP57691, ABP57692, ABP57693, ABP57694,
DR   ABP57695, ABP57696, ABP57697, ABP57698, ABP57699, ABP57700,
DR   ABP57701, ABP57702, ABP57703, ABP57704.
XX
PT   New butenyl-spinosyn biosynthetic genes, useful for increasing the
PT   production of butenyl-spinosyn insecticidal macrolides, or for changing
PT   the metabolites or products produced by spinosyn-producing
PT   microorganisms  -
XX
PS   Claim 2; Page 99-119; 218pp; English.
XX
CC   The invention relates to a novel DNA molecule comprising a DNA sequence
CC   that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
CC   PKS (polyketide synthetase) domain, or a spinosyn PKS module. The
CC   butenyl-spinosyn biosynthetic genes are useful for increasing the
CC   production of butenyl-spinosyn insecticidal macrolides. The genes are
CC   also useful for changing the metabolites or products produced by
CC   spinosyn-producing microorganisms. The present sequence represents a DNA
CC   molecule encoding butenyl-spinosyn biosynthetic enzymes.
XX
SQ   Sequence 36538 BP; 6867 A; 12266 C; 11182 G; 6223 T; 0 other;

  Query Match             9.5%;  Score 119.6;  DB 25;  Length 36538;
  Best Local Similarity   48.2%;  Pred. No. 9.6e-18;
  Matches  424;  Conservative    0;  Mismatches  429;  Indels   27;  Gaps    2;

Qy        291 GCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACG 350
              | ||   |||  ||| ||   | ||| | ||||| || |||      | |||| ||   ||||
Db      24626 GATGGACCCGCCGGAGCATACGCGGCTGCGCAAACTGGTCGGTAAGGCGTTCACCGCACG 24567

Qy        351 CGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCG 410
                  |||| | ||| ||| ||    |  ||||     ||| ||||| || || |||
Db      24566 TCGAGTCGAACAGCTTCGCCCCCGCACCCAGCAGCTCGTTGACCACCTACTTGATCGGAT 24507

Qy        411 CTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGC 470
```

```
Db      24506 GGCCGCGGACGGGCCGCCTGGCGACCTGGTGTCAGCCCTGGCCCTGCCGCTGCCCATCAA 24447

Qy        471 GATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTC 530

Db      24446 GGTGATCTGCGACTTGCTGGGCATTCCGGTGGCCGACCGCGAGCGCTTCCGGGTCTGGTC 24387

Qy        531 GGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGAC 590

Db      24386 GGACATTGCGTTGGCGATCACATCGAACAGTCCTGAGGAGATTCGGGAAAGCCGCGATCA 24327

Qy        591 CCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCG 650

Db      24326 AATTCGCGCGTATATCGGAGAACTGGTCC-----------------------AGCAGCG 24291

Qy        651 CAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAG 710

Db      24290 CAAGAAGATGCCGACTGAGGACCTGCTGAGCGTGCTGGTCCAGGCCAGAGCGGAAGGAGC 24231

Qy        711 CAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGA 770

Db      24230 CCAGCTGAGCGAGGAGGAGATTGTCGTCACCGGCGCGGGTCTGCTGATCGCCGGGTTCGA 24171

Qy        771 TACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCT 830

Db      24170 GACCACCGCTAACCACATCGCAAACTTCACCTTCAACCTGCTCACCCATCCGGACCAGCT 24111

Qy        831 CGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTT 890

Db      24110 TGACAAGCTGATCGCCGATCCAGAGCTGGTGCCGCGCGCGGTGGAGGAGTTGCTGCGCTA 24051

Qy        891 CGACAATATCCT---CAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTG 947

Db      24050 CACGCCGCTCGGTGCCACACCTGGGTTCCCGCGCATCGCGACTGAGGATCTCGAACTGGG 23991

Qy        948 CGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGA 1007

Db      23990 CGGAGTGTCGATTCGCCGGGGTGACGCAGTGTTCTTCGAGATAGCCTCGGCGAATCGTGA 23931

Qy       1008 TGGGACTGTATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCT 1067

Db      23930 CAGTGCCGTCTTCGACGGGCCGGACGAACTCGACCTGGCTCGGGAGCACAACTCGCACAT 23871

Qy       1068 CGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGA 1127

Db      23870 GGCGCTGGGGCACGGCCCGCACTACTGCATCGGCGCCCAATTGGCGAGGATGGAGCTGCA 23811

Qy       1128 GATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGATG 1167

Db      23810 GGTAGCGATCGGCACGCTGATCAAACGGTTTCCGCAGCTG 23771
```

```
RESULT 14
ABN88911
ID   ABN88911 standard; DNA; 1449 BP.
XX
AC   ABN88911;
XX
DT   21-AUG-2002  (first entry)
XX
DE   Streptomyces sp. cytochrome P450 encoding DNA SEQ ID NO:3.
XX
KW   Streptomyces; cytochrome P450; anticancer; UCN-01; monooxygenase;
KW   microorganism; gene; ds.
XX
OS   Streptomyces sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1449
FT                   /*tag=  a
FT                   /product= "cytochrome P450"
XX
PN   JP2002058490-A.
XX
PD   26-FEB-2002.
XX
PF   22-AUG-2000; 2000JP-0251304.
XX
PR   22-AUG-2000; 2000JP-0251304.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
DR   WPI; 2002-410015/44.
DR   P-PSDB; ABB81316.
XX
PT   Novel cytochrome P450 used for improving the productivity of anticancer
PT   drug UCN-01  -
XX
PS   Claim 5; Page 12-14; 22pp; Japanese.
XX
CC   The present invention describes 3 Streptomyces cytochrome P450 (I)
CC   sequences, ABB81315 to ABB81317 encoded by ABN88910 to ABN88912
CC   respectively. (I) have monooxygenase activity and improve the
CC   productivity of an anticancer drug designated UCN-01. (I) is used
CC   for the production of UCN-01 with transformed Streptomyces sp.
CC   microorganisms.
XX
SQ   Sequence 1449 BP; 231 A; 620 C; 390 G; 208 T; 0 other;

  Query Match             9.4%;  Score 118.6;  DB 24;  Length 1449;
  Best Local Similarity   48.5%;  Pred. No. 9.3e-18;
  Matches  417;  Conservative    0;  Mismatches  409;  Indels   33;  Gaps    2;

Qy        305 ATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGC 364

Db        521 AACACGCCCGGCTACGGGCGCTGGTGACGAAGGCGTTCACCACGGGGGCGGTGGCGCGTC 580

Qy        365 TGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGG 424

Db        581 TGCGCCCCTACATCTCCTCCCTGGTCGACGAGTTGCTCGACACCTGGCCGACCCACGGAA 640

Qy        425 AGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGT 484

Db        641 CGGTGGACCTGATCGCCGACCTCGCGGTGCCGCTCCCCGTCACGGTCATCTGCGAGCTGC 700

Qy        485 TGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCG 544

Db        701 TCGGGGTGCCGGATTC--------------------------------CGACCGCGCGTC 728

Qy        545 CGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCG 604

Db        729 CGTACGCACCTGGTCGAGCG-ACCTGTTCGCTGCCGGAGACCCGCAGCGGATCGACGCCG 787

Qy        605 TCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCG 664

Db        788 CCTCCCACGCCGTCGGCGACTACATGACCGCCCTCGTCGCCGCCAAGCGCACCGCACCCG 847

Qy        665 AAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGA 724

Db        848 GCGACAGCCTGCTCGACGACCTCATCGCCGTACGCGACGGCCAGGACCACCTGTCCGAGG 907

Qy        725 AGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACC 784

Db        908 ACGAACTCGTATCGCTCGCCGTACTCCTGCTGGTGGCCGGCCACGAGACCACCACCAACT 967

Qy        785 TTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGG 844

Db        968 TCATCGGCAACGCCGCCCTGGCCCTCCTACGGCACCCGGAGTCGCTCGCGCACCTGAGGG 1027

Qy        845 CCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCA 904

Db       1028 CCGAGCCCGCAACTCCTCGGCGGCGCACTGGACGAGTTGCTGCGCTACGACTCGCCCGTCG 1087

Qy        905 GAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGA 964

Db       1088 GGATCGCGACGTTCCGCTTCAGCACCGAGGCGCTCACGCTCGGCGGCACCGAGATCCCCG 1147
```

```
Qy        965 AAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGACTGTATTCTCCA 1024
Db       1148 AAGGCGTACCGGTCCTCATCGCACCCGGCGCCGCCAATCGCGACCCGGACCGCTTCCCCG 1207

Qy       1025 GGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCC 1084
Db       1208 ACCCCGACCGCCTGGACCTCACCCGCGGCGCCACCGGTCACCTGGCCTTCGGCCACGGCA 1267

Qy       1085 CCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCACCA 1144
Db       1268 TCCACCGCTGCCTGGGAGCCCCGTTGGCCCGAGCCGAAGCCGAACTGGCCCTCCACGCCG 1327

Qy       1145 TCTTCCGTAGGTTCCCCGA 1163
Db       1328 TCATCACCCGCTACCCCCA 1346
```

```
RESULT 15
ABX56043
ID   ABX56043 standard; DNA; 1209 BP.
XX
AC   ABX56043;
XX
DT   13-FEB-2003  (first entry)
XX
DE   M. echinospora calicheamicin biosynthesis gene calS.
XX
KW   Calicheamicin biosynthetic gene cluster; aryltetrasaccharide;
KW   aglycone; calicheamicin resistance; nonchromoprotein enediyne;
KW   enediyne resistance; bone marrow cell; gene therapy; gene; ds.
XX
OS   Micromonospora echinospora spp. calichensis.
XX
PN   WO200279465-A2.
XX
PD   10-OCT-2002.
XX
PF   28-NOV-2001; 2001WO-US44285.
XX
PR   28-NOV-2000; 2000US-0724797.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Thorson J;
XX
DR   WPI; 2003-092897/08.
DR   P-PSDB; ABU11981.
XX
PT   Novel nucleic acid molecule from nonchromoprotein enediyne biosynthetic
PT   gene cluster from Micromonospora echinospora useful for conferring
PT   calicheamicin resistance on a subject -
XX
PS   Claim 9; Page 110-111; 179pp; English.
XX
CC   The present invention relates to the isolation of the Micromonospora
CC   echinospora spp. calcichensis calicheamicin biosynthetic gene cluster
CC   encoding proteins and enzymes used in the biosynthetic production of
CC   calicheamicin, including aryltetrasaccharide and aglycone. The gene
CC   cluster also includes the gene encoding for the protein that confers
CC   calicheamicin resistance. The calicheamicin biosynthetic gene cluster
CC   is a nonchromoprotein enediyne biosynthetic gene cluster. Expression
CC   vectors containing genes from the biosynthetic gene cluster are also
CC   disclosed. The expression vectors are useful for producing calicheamicin
CC   biosynthetic proteins. The calicheamicin self-resistance gene provides
CC   an approach for gene therapy, for example, by introduction of enediyne
CC   resistance genes into bone marrow cells, thus increasing resistance and
CC   allowing tolerance to chemotherapeutic doses of calicheamicin.
CC   ABX56026-ABX56073 represent genes from the M. echinospora calicheamicin
CC   biosynthesis gene cluster.
XX
SQ   Sequence 1209 BP; 132 A; 451 C; 452 G; 174 T; 0 other;
```

```
  Query Match             9.3%;  Score 116.8;  DB 25;  Length 1209;
  Best Local Similarity   47.7%;  Pred. No. 2.3e-17;
  Matches  488;  Conservative    0;  Mismatches  512;  Indels   24;  Gaps    4;

Qy        128 ACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCGGTGTTCC 187
Db        101 ACCGTGGCGCCCACCGGGTCTGGTACGTCTCCCGCTTCGCGGACGTGCGCGCGGTGCTCG 160

Qy        188 GCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTACTCGTCGG 247
Db        161 GCGACGAGCGCTTCGCCCGGACCGGCATCCGCCGGTTCTGGACCGACCTCGTCGGGCCCG 220

Qy        248 CCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCGGAGGATC 307
Db        221 GGCT------GCTCGCCGAGATCGTCGGCGACATCATCCTGTTCCAGGACGAGCCCGACC 274

Qy        308 ACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGC 367
Db        275 ACGGCCGGCTGCGCGGGGTGGTCGGCCCGGCGTTCTCGCCGTCCGCGCTGCGCCGGCTGG 334

Qy        368 GCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGT 427
Db        335 AACCGGTGATCGCCGGCACCGTGGACGACCTGCTGCGGCCCGCCCTGGCCCGGGGCGCGA 394

Qy        428 TCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGA 487
Db        395 TGGACGTGGTCGACGAGCTGGCGTACCCGCTGGCGCTGCGCGCGGTCCTCGGCCTGCTCG 454

Qy        488 AGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGC 547
Db        455 GCCTGCCCGCCGCCGACTGGGGGGCGGTCGGGCGCT-----------GGTCGCGCGACG 502

Qy        548 TCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCA 607
Db        503 TGGGACGGACCCTGGACCGGGGCGCCAGCGCCGAGGACATGCGCCCGCGGCCACGCGGCGA 562

Qy        608 CCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAA 667
Db        563 TCGCCGAGTTCGCCGACTACGTGGAGCGGGCCCTCGCGAGGCGGCGGCGTGAGGGCGGCG 622

Qy        668 ATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGG 727
Db        623 AGGACCTGCTGGCGTTGATGCTCGACGCCCA---CGACCGCGGCCTGATGAGTCGCAACG 679

Qy        728 AGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTA 787
Db        680 AGATCGTCAGCACGGTGGTCACGTTCATCTTCACCGGCCACGAGACGGTGGCCAGCCAGG 739

Qy        788 TCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCG 847
Db        740 TGGGCAACGCCGTGCTGAGCCTGCTGGCGCACCCGGACCAGCTCGACCTGCTCCGGCGCC 799

Qy        848 AGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAA 907
Db        800 GGCCGGACCTGCTGGCCCAGGCCGTCGAGGAGTGCCTGCGGTACGACCCGTCGGTGCAGT 859

Qy        908 TAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGAAAG 967
Db        860 CCAACACCCGGCAGCTCGAC---GTCGACGTGGAGCTGCGCGGTCGGCGGCTGCGCCGCG 916

Qy        968 GGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGACTGTATTCTCCAGGC 1027
Db        917 ACGACGTCGTGGTCGTCCTGGCCGGCGCGGCGAACCGGGACCCGCGGCGGTACGACCGGC 976

Qy       1028 CAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCCCC 1087
Db        977 CCGACGATTTCGACATCGAGCGGGATCCGGTCCCGTCGATGTCCTTCGGCGCGGGGATGC 1036

Qy       1088 ATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCACCATCT 1147
Db       1037 GCTACTGCCTCGGGTCCTACCTCGCCCGTACGCAGCTGCGCGCCGCGGTGGCCGCCCTGG 1096

Qy       1148 TCCG 1151
```

```
             |||
Db       1097 CCCG 1100
```

Search completed: October 3, 2003, 12:28:58
Job time : 317.318 secs

```
OM nucleic - nucleic search, using sw model

Run on:         October  3, 2003, 11:40:41 ; Search time 4507.41 Seconds
                                             (without alignments)
                                             11435.853 Million cell updates/sec

Title:          US-10-014-717-1_COPY_62369_63628
Perfect score:  1260
Sequence:       1 atgacacaggagcaagcgaa..........agccctccaaagctggatag 1260

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters:      5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1260	100.0	16124	6	AX024277	AX024277 Sequence
	2	1260	100.0	68750	1	AF210843	AF210843 Sorangium
	3	1260	100.0	68750	6	AR193029	AR193029 Sequence
	4	1260	100.0	68750	6	AR199551	AR199551 Sequence
	5	1260	100.0	68750	6	AR199559	AR199559 Sequence
	6	1260	100.0	68750	6	AR199567	AR199567 Sequence
	7	1260	100.0	68750	6	AR201097	AR201097 Sequence
	8	1260	100.0	68750	6	AR208671	AR208671 Sequence
	9	1246.2	98.9	58733	1	AF217189	AF217189 Sorangium
	10	1246.2	98.9	71989	6	AR172664	AR172664 Sequence
c	11	157.4	12.5	302650	1	AP005958	AP005958 Bradyrhiz
	12	154	12.2	3078	1	SERCP450A	M83110 Saccharopol
	13	149.4	11.9	1233	6	AX697983	AX697983 Sequence
c	14	149.4	11.9	60196	6	AX697977	AX697977 Sequence
c	15	149.2	11.8	66808	1	SAU421825	AJ421825 Stigmatel
	16	148.8	11.8	301925	1	AP005046	AP005046 Streptomy
	17	146.4	11.6	78210	1	AB070949	AB070949 Streptomy
c	18	146.4	11.6	300425	1	AP005022	AP005022 Streptomy
	19	143.8	11.4	1257	6	AX097457	AX097457 Sequence
	20	143.8	11.4	14759	1	AE004755	AE004755 Pseudomon
	21	141.8	11.3	10167	1	SFU08223	U08223 Streptomyce
	22	132	10.5	24268	1	AB070947	AB070947 Streptomy
c	23	132	10.5	300550	1	AP005030	AP005030 Streptomy
	24	131.6	10.4	12420	1	AF521896S2	AF521895 Streptomy
	25	130	10.3	2243	1	SERERYFGH	M54983 Saccharopol
	26	128.8	10.2	1347	6	AX573690	AX573690 Sequence
	27	128.8	10.2	90348	1	AF497482	AF497482 Micromono
c	28	127	10.1	298550	1	AP005029	AP005029 Streptomy
	29	123.6	9.8	17512	1	AB071405	AB071405 Lechevali
	30	123.6	9.8	25681	1	SAE414559	AJ414559 Saccharot
	31	123.6	9.8	26144	1	AB090952	AB090952 Lechevali
c	32	123.6	9.8	28654	1	AF534707	AF534707 Lechevali
	33	123.4	9.8	325483	1	AP005050	AP005050 Streptomy
	34	123	9.8	1470	1	AF087022	AF087022 Streptomy
	35	123	9.8	4342	1	AF079139	AF079139 Streptomy
	36	123	9.8	5970	6	AR271682	AR271682 Sequence
	37	123	9.8	5970	6	AR277666	AR277666 Sequence
	38	123	9.8	32634	1	AF293355	AF293355 Streptomy
	39	122.6	9.7	1191	6	BD133544	BD133544 Novel cyt
	40	121.8	9.7	299925	1	AP005039	AP005039 Streptomy
c	41	121.2	9.6	69644	1	AY179507	AY179507 Streptomy
c	42	120.6	9.6	210614	1	AB088224	AB088224 Streptomy
	43	120.4	9.6	6629	1	AB072568	AB072568 Streptomy
c	44	120.4	9.6	19601	1	AB070944	AB070944 Streptomy
c	45	120.4	9.6	299425	1	AP005049	AP005049 Streptomy

ALIGNMENTS

```
RESULT 1
AX024277
LOCUS       AX024277               16124 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024277 AX024384
VERSION     AX024277.1  GI:10184551
KEYWORDS    .
SOURCE      Polyangium cellulosum
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 82 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
```

```
COMMENT     On Oct 15, 2002 this sequence version replaced gi:10184588.
FEATURES             Location/Qualifiers
     source          1..16124
                     /organism="Polyangium cellulosum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match             100.0%;  Score 1260;  DB 6;  Length 16124;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-199;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7086 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 7145

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7146 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 7205

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7206 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 7265

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7266 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 7325

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7326 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 7385

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7386 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 7445

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7446 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 7505

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7506 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 7565

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7566 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 7625

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7626 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 7685

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7686 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 7745

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7746 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 7805

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7806 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 7865

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7866 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 7925

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7926 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 7985

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7986 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 8045

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8046 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 8105

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8106 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 8165

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8166 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 8225

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8226 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 8285

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8286 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 8345


RESULT 2
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="Polyangium cellulosum"
                     /mol_type="genomic DNA"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1..1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
```

```
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900..3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
                     /db_xref="GI:6724239"
                     /translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
                     AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
                     EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
                     SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
                     VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
                     GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
                     TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
                     IGIVMTQSAGFLFSGALERFWRSVYVATESA"
     CDS             3415..5556
                     /note="ORF3; similar to Na/H and drug/H antiporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26906.1"
                     /db_xref="GI:6724240"
                     /translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
                     GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
                     RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
                     RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
                     LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
                     NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
                     LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
                     SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
                     IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
                     IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
                     ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
                     ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
                     YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
     CDS             complement(5612..5992)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26907.1"
                     /db_xref="GI:6724241"
                     /translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTIAYTDRE"
     CDS             6226..6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MKHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYANFELLITAEKQSSPQSPSSAAVIGPTSVG"
     gene            7610..11875
                     /gene="epoA"
     CDS             7610..11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
     misc_feature    7640..8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
     misc_feature    9236..10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
     misc_feature    10529..11428
                     /gene="epoA"
                     /note="Region: enoyl reductase"
     misc_feature    11549..11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
     gene            11872..16104
                     /gene="epoP"
     CDS             11872..16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
     gene            16251..21749
```

```
     CDS              /gene="epoB"
                      16251. .21749
                      /gene="epoB"
                      /note="EPOS B"
                      /codon_start=1
                      /transl_table=11
                      /product="polyketide synthase"
                      /protein_id="AAF26920.1"

  Query Match             100.0%;  Score 1260;  DB 1;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-199;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628


RESULT 3
AR193029
LOCUS       AR193029            68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 1260;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-199;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728
```

```
Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628

RESULT 4
AR199551
LOCUS       AR199551                 68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTORS    Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 1260;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-199;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268
```

```
Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGACTGTATTC 63388

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628
```

```
RESULT 5
AR199559
LOCUS       AR199559                68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn.. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 1260;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-199;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGACTGTATTC 63388

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628
```

```
RESULT 6
AR199567
LOCUS       AR199567                68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
```

```
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 1260;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-199;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628


RESULT 7
AR201097
LOCUS       AR201097                68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 1260;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-199;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628

RESULT 8
AR208671
LOCUS       AR208671                68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 1260;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-199;
  Matches 1260;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62369 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 62428

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62429 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 62488

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62489 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 62548

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62549 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 62608

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 62668

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62669 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 62728

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62729 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 62788

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62789 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 62848

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62849 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 62908

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62909 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 62968

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62969 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 63028

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63029 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 63088

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63089 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 63148

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63149 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 63208

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63209 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 63268

Qy         901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63269 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 63328

Qy         961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63329 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 63388

Qy        1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63389 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 63448

Qy        1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63449 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 63508

Qy        1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63509 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 63568

Qy        1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       63569 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628


RESULT 9
AF217189
LOCUS       AF217189               58733 bp    DNA     linear   BCT 09-JUN-2000
DEFINITION  Sorangium cellulosum putative transposase gene, partial cds;
            putative transposase gene, complete cds; epothilone biosynthesis
            gene cluster, complete sequence; putative membrane protein gene,
            complete cds.
ACCESSION   AF217189
VERSION     AF217189.1  GI:7453554
KEYWORDS    .
SOURCE      Polyangium cellulosum
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 58733)
  AUTHORS   Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and
            Julien,B.
  TITLE     Cloning and heterologous expression of the epothilone gene cluster
  JOURNAL   Science 287 (5453), 640-642 (2000)
  MEDLINE   20115953
   PUBMED   10649995
REFERENCE   2  (bases 1 to 58733)
  AUTHORS   Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
  TITLE     Isolation and characterization of the epothilone biosynthetic gene
            cluster from Sorangium cellulosum
  JOURNAL   Gene 249 (1-2), 153-160 (2000)
  MEDLINE   20293058
   PUBMED   10831849
REFERENCE   3  (bases 1 to 58733)
  AUTHORS   Julien,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
     source          1. .58733
                     /organism="Polyangium cellulosum"
                     /mol_type="genomic DNA"
                     /strain="SMP44"
                     /db_xref="taxon:56"
     CDS             <1. .992
                     /note="orfA"
                     /codon_start=3
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62878.1"
                     /db_xref="GI:7453555"
                     /translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
                     SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
                     DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
                     DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
                     GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
                     HIGAVRQVIDRRRSERHLPPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"
     CDS             989. .1501
                     /note="orfB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62879.1"
                     /db_xref="GI:7453556"
                     /translation="MTDLTPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
                     RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
                     AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
                     RTKQRRTKKH"
     gene            1998. .6263
                     /gene="epoA"
     CDS             1998. .6263
                     /gene="epoA"
                     /note="type I polyketide synthase; epothilone PKS loading
                     module"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoA"
                     /protein_id="AAF62880.1"
                     /db_xref="GI:7453557"
                     /translation="MADRPIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
                     GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
                     PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
                     DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
                     QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"
     gene            6260. .10492
                     /gene="epoB"
     CDS             6260. .10492
                     /gene="epoB"
                     /note="non ribosomal peptide synthetase; epothilone PKS
                     module 1"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoB"
                     /protein_id="AAF62881.1"
                     /db_xref="GI:7453558"
                     /translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
                     PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
```

```
                QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
                ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
                GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
                GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
                EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
                GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
                STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
                FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
                PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
                FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"
gene            10639. .16137
                /gene="epoC"
CDS             10639. .16137
                /gene="epoC"
                /note="type I polyketide synthase; epothilone PKS module
                2"
                /codon_start=1
                /transl_table=11
                /product="EpoC"
                /protein_id="AAF62882.1"
                /db_xref="GI:7453559"
                /translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
                ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
                ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
                HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
                AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
                GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
                DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
                ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
                FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
                REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
                SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
                VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
                GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
                EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
                GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
                VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST
                QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
gene            16134. .37907
                /gene="epoD"
CDS             16134. .37907
                /gene="epoD"
                /note="type I polyketide synthase; epothilone PKS modules
                3, 4, 5, and 6"
                /codon_start=1
                /transl_table=11
                /product="epoD"
                /protein_id="AAF62883.1"
                /db_xref="GI:7453560"
                /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
```

```
  Query Match             98.9%;  Score 1246.2;  DB 1;  Length 58733;
  Best Local Similarity   99.4%;  Pred. No. 3.5e-197;
  Matches 1251;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56757 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 56816

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db      56817 GCGCCTGGGTACGCGGAGGACCCGTTTCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 56876

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56877 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 56936

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      56937 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAATGGGAATCGAGCGCGGAGTAC 56996

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56997 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 57056

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||| ||||||
Db      57057 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCATCGTTTACGTCACGCGCGATCGAC 57116

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57117 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 57176

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      57177 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGTGCGATCAGCGCT 57236

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57237 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 57296

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      57297 CGCGCGCTCGGCGTGGGTTTGGTGCCCCGGGTCGATGAGGAGACCAAGACCCTGGTCGCG 57356

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      57357 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGGCGTCCTCGATGAGCGGCGCAGGAACCCG 57416

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57417 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 57476

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57477 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 57536

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57537 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 57596

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57597 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 57656

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57657 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 57716

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57717 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 57776

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      57777 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGAGCGCGAGCCTCGCGTACGGTAGA 57836

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57837 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 57896

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57897 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 57956

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATA 1259
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57957 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATA 58015


RESULT 10
AR172664
LOCUS       AR172664                71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1..71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others
ORIGIN

  Query Match            98.9%;  Score 1246.2;  DB 6;  Length 71989;
  Best Local Similarity  99.4%;  Pred. No. 3.4e-197;
  Matches 1251;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56757 ATGACACAGGAGCAAGCGAATCAGAGTGAGACGAAGCCTGCTTTCGACTTCAAGCCGTTC 56816

Qy         61 GCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 120
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db      56817 GCGCCTGGGTACGCGGAGGACCCGTTTCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCC 56876

Qy        121 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56877 ATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCACGACGTGTCGGCG 56936

Qy        181 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCGAGCGCGGAGTAC 240
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      56937 GTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAATGGGAATCGAGCGCGGAGTAC 56996

Qy        241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56997 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTGTTCGGGCTGCCGCCG 57056

Qy        301 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATCGAC 360
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||| ||||||
Db      57057 GAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCATCGTTTACGTCACGCGCGATCGAC 57116

Qy        361 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57117 CTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAA 57176

Qy        421 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      57177 GAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGTGCGATCAGCGCT 57236

Qy        481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57237 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 57296

Qy        541 CGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTCGCG 600
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      57297 CGCGCGCTCGGCGTGGGTTTGGTGCCCCGGGTCGATGAGGAGACCAAGACCCTGGTCGCG 57356

Qy        601 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAACCCG 660
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      57357 TCCGTCACCGAGGGGCTCGCGCTGCTCCATGGCGTCCTCGATGAGCGGCGCAGGAACCCG 57416

Qy        661 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57417 CTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGC 57476

Qy        721 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57477 ACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACGATC 57536

Qy        781 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57537 TACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTG 57596

Qy        841 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57597 AAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAATATC 57656

Qy        901 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57657 CTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCGATC 57716

Qy        961 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57717 AAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTATTC 57776

Qy       1021 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGA 1080
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      57777 TCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGAGCGCGAGCCTCGCGTACGGTAGA 57836

Qy       1081 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57837 GGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGC 57896

Qy       1141 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57897 ACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATACCAC 57956

Qy       1201 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATA 1259
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57957 CCCGCGTTCCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATA 58015


RESULT 11
AP005958/c
LOCUS       AP005958               302650 bp    DNA     linear   BCT 28-MAR-2003
DEFINITION  Bradyrhizobium japonicum USDA 110 DNA, complete genome, section
            24/31.
ACCESSION   AP005958 BA000040
VERSION     AP005958.1  GI:27354550
KEYWORDS    .
SOURCE      Bradyrhizobium japonicum USDA 110
```

```
  ORGANISM  Bradyrhizobium japonicum USDA 110
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE   1
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,
            Sasamoto,S., Watanabe,A., Idesawa,K., Iriguchi,M., Kawashima,K.,
            Kohara,M., Matsumoto,M., Shimpo,S., Tsuruoka,H., Wada,T., Yamada,M.
            and Tabata,S.
  TITLE     Complete genomic sequence of nitrogen-fixing symbiotic bacterium
            Bradyrhizobium japonicum USDA110
  JOURNAL   DNA Res. 9 (6), 189-197 (2002)
  MEDLINE   22484998
   PUBMED   12597275
REFERENCE   2
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,
            Sasamoto,S., Watanabe,A., Idesawa,K., Iriguchi,M., Kawashima,K.,
            Kohara,M., Matsumoto,M., Shimpo,S., Tsuruoka,H., Wada,T., Yamada,M.
            and Tabata,S.
  TITLE     Complete genomic sequence of nitrogen-fixing symbiotic bacterium
            Bradyrhizobium japonicum USDA110 (supplement)
  JOURNAL   DNA Res. 9 (6), 225-256 (2002)
  MEDLINE   22485002
   PUBMED   12597279
REFERENCE   3  (bases 1 to 302650)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1..302650
                     /organism="Bradyrhizobium japonicum USDA 110"
                     /mol_type="genomic DNA"
                     /strain="USDA110"
                     /db_xref="taxon:224911"
     gene            31..1110
                     /gene="blr6272"
     CDS             31..1110
                     /gene="blr6272"
                     /note="AraC family
                     ORF_ID:blr6272"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulatory protein"
                     /protein_id="BAC51537.1"
                     /db_xref="GI:27354551"
                     /translation="MENLLESLVESFGLGGLPVAASGRDAKRGRAMFTDLQAAPAWID
                     LQAEPRDHQLRQAAQAIPRERGWRRIDRGPEAAIEDITISRWEDFRDSSSHDATTPED
                     RYFVGIALKATRAKLTRDRQVIFDGTMPAGTLYVSAPSKPLAAQFQAPCAFLHFHISA
                     DHFPAQSPAAEGLGDLVLLRDPLAAELARALTEHGDAADREFTRCVGQTLAMHLARLE
                     PPRARVNALPKWRLRRVEQYIADNFDRCISLSELANVAGLSRMHFAAQFRTATGYRPR
                     EYLLNHRIEHAKTLLATTGRPLAEIALAVGFSTQAHFSTVFKRISGQSPARWRLASKA
                     EPLAAEALPRRRPAADRDWMASAAA"
     gene            complement(1428..3794)
                     /gene="bll6273"
     CDS             complement(1428..3794)
                     /gene="bll6273"
                     /note="ORF_ID:bll6273
                     probable biotin sulfoxide reductase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC51538.1"
                     /db_xref="GI:27354552"
                     /translation="MDRYTATMYRNKDQMMDDTIGFPDPGLDLSDGFKPHTSHWGVFS
                     ARNGAAGLEVRAYAGDPDPNGIIDNFPGALRHQARIAQPAIRRGWLERGPGPDDRRGR
                     DEFVSVSWEKALDLLGDELGRIRDTRGPGAVFGGSYGWSSAGRFHHAQSQVHRFLNIA
                     MGGYVRSVNSYSSGASSVLLPQILAGYEDITKRNVTWEQIATETDIVLAFGGMALKNS
                     MVAGGSISKHVERSAMEAARKRGCEFILVSPLREDLPVEAGAEWMTCVPSTDTALMLG
                     IVHTLVSEKLHDQAFLDRYTEGWPVFLRYLTGESDGQPKSAEWAAAISGVDADTIRTL
                     ARRLAGKRALITVSHSLQRAEHGEQPVWMGMVLAAALGQIGLPGGGYAYSLGAIGYYG
                     RRVNDVPGPTLGQGRNGVRDFIPVARIADMLLNPGATYRYNGETRTYPDIRLVYWAGG
                     NPFHHHQDINRLRKAFAQLDTFVVHELAWTATARHADIVLPSTMTLEREDIGYSSNDP
                     LMVAMHRIAEPFGLARDDYEIFADLAERLGAREPFTEGRTSREWLKHLYEPTRASLAM
                     RGLEAPSFDEFWQRGSLVVPQQPDDGGRLRRFREDPVAHALPTPSGRIEIFSTRIAGH
                     GDADCPGHPVWLDKTDMPKPGAPCFLVANQPVTRLHSQLDFGGHSLGAKHRGREVARM
                     NPVDADARGIKDGDIIRLFNERGACLAAVHVTDGIAPGVVQLPTGAWYDPMDPEDDAP
                     LCVHGNPNVLTRDVGTSSFAQGCTGQLTTVEVEKFTGNLPPIRAFDPV"
     gene            complement(3898..4947)
                     /gene="bll6274"
     CDS             complement(3898..4947)
                     /gene="bll6274"
                     /note="ORF_ID:bll6274
                     putative dehydrogenase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC51539.1"
                     /db_xref="GI:27354553"
                     /translation="MTGASDTVAIGADALRPFAADLLRAGGFTADQADETAAMLVWAN
                     LRGADSHGVLRIPRYVEMVELGLINPAAEPRQVSGKGAVAVIDAQRAPGAVGMSLAAR
                     KASELADTHGIGWCAVRSISHAGAIGYYAHQVAQAGRIGIVMTASKPLMIYHGSSAEG
                     VSTNPIAIAAPTGNPDRPLLLDMSTASVALGKIMAARDAGQPIPVGWAVDGNGQATTD
                     PREAKAVLPMAGPKGSGLSLMIEVLASVLTANPLIAVALVHGGDPGGNGLVAVLDPSA
                     FGDNFTLEVDRLGAAIKSLPPAAGADSVYLPGERGFEEMERRSRAGIPLARGTVSRLL
                     ALASKLNVAVPEGVA"
     gene            complement(4950..5873)
                     /gene="bll6275"
     CDS             complement(4950..5873)
                     /gene="bll6275"
                     /note="ORF_ID:bll6275
                     hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC51540.1"
                     /db_xref="GI:27354554"
                     /translation="MSAPPLSQPTREIPMQEKHTFRSDGLKLSAVLHVPDARKPGQRL
                     PAFIVLHGFVGSKDESHAEIQARMLEQMGYVALRFDFRCCGESEGERAQVRCFDQVAD
                     AKNALTFLAEREEVDPGRIGVVGHSFGAAVSVYAAGVDDRFACVISSCGWGDGERKFR
                     GQHPTPEAWAKFTSLLEKGRKQKQETGQSMWISRFDAVPIPEHLRKNLSPKAIMEIPV
                     ETAWSMYNFRADEVVGNIAPRPILFLHTANDTITPTEQSIRMFEKAGQPAELVLITGT
                     SHFPLAEGDAPRTKAIIKGWLDKFFPSPLGA"
     gene            complement(5910..6812)
                     /gene="bll6276"
     CDS             complement(5910..6812)
                     /gene="bll6276"
                     /note="ORF_ID:bll6276
                     unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC51541.1"
                     /db_xref="GI:27354555"
                     /translation="MPPIVHAVGHVKINTTAIEAVVRDATEILGLQITYTDEGQTWLS
                     SNGRAAELVLLRSHENATHTIGLEALTVDAVREAESRVESAGCRVISRKPSLDCMAAG
                     VTFDTPEGLRFEVHTPTRDQIYGRRYPTHGVGPNRMDHLNLTSPDPIATRQQLERICG
                     LRISERMVDDSLSWMYGGNRQHHIVGVVRGKVGLHHYSFEFKEFNDYLKLGDLLDRFD
                     RQMLWGPGRHRPGDNTYAYYTDSSGAMIECSGSMALIADDDGFVPNVITNLERPGNVR
                     CMNVWGTPAPLEWRQFHFPFTAIT"
     gene            7080..7853
                     /gene="blr6277"
     CDS             7080..7853
                     /gene="blr6277"
                     /note="GntR family
                     ORF_ID:blr6277"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulatory protein"
                     /protein_id="BAC51542.1"
                     /db_xref="GI:27354556"
                     /translation="MILTGSRGMAGRRRMRERVKPAGETRAADVLHRMRADIISCTLK
                     PGAKLRFEALRDIYAVSFSTLREALSRLVAEGLVIAEDQRGFVVAPVSIDDLNDLTYV
                     RVLIERECVALAIKNGDDAWEADIIGAFHRMDRLQNRLGSNYYLSEEWGKLHGDFHFS
                     LVAACGSPNLLEIRQKLFERAHRYRRMSSQFRTKWRAKDVEHKMIMDSVVARDAAKAE
                     ELIERHIRETTDNVIKHAGHLFVTSDEERPRRTSRVAAE"
     gene            complement(7993..8880)
```

```
                     /gene="ilvE"
     CDS             complement(7993..8880)
                     /gene="ilvE"
                     /note="ORF_ID:bll6278"
                     /codon_start=1
                     /transl_table=11
                     /product="branched-chain-amino-acid transaminase"
                     /protein_id="BAC51543.1"
                     /db_xref="GI:27354557"
                     /translation="MVPDTNSISHSRTWTFFRGQWHEGNVPIMGPRTHGLWLASTIFD
                     GARAFEGVMPDLDRHCARVNQSAVNFGLKPVVDLDTWLGLVREGIARFPVNAELYIRP
                     MYWPQNGIGGGVLFDPETTDWCLCIYESPIPEPVGSAITLSPFRRPTAECAPADVKAS
                     CLYPNSSRALREAASRGFQNCLMLDMLGNVTEFGNSNVFMAKDGVVFTPAPNGTFLNG
                     ITRQRVIELLRGDGITVLEATLRYSDFVSADEIFSTGNFAKVAPVIRIDDRQLRPGRF
                     YARARKLYWDFAHGMKLAA"
     gene            complement(8909..9784)
                     /gene="bll6279"
     CDS             complement(8909..9784)
                     /gene="bll6279"
                     /note="ORF_ID:bll6279
                     hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC51544.1"
                     /db_xref="GI:27354558"
                     /translation="MRVRTMIRCREIGIADLDAIADLLTRGFPTRSRDYWIDGLRRLS
                     VRETPDSFPRFGYMLEHDRAPVGALQLIHTARNDGGSISIRCNLSSWYVEPAFRSYAP
                     MLTKVAQRHKDVTYVNISPAEWTWQVIEGQGFRRYCGGLFFSFPALSRPAKAMRVEVV
                     RQDAQAIDGLSEADVVLLARHAAYGCLCLICRASDGRAFPFVLQPMRVRRIRLLAMQM
                     IYCPDVADYVVCAGAIGRFLLRRGRISIALDANGPVRGLVGIYREPIGRKYFKGPHCP
                     RLADLSDSELVIYGP"
     gene            complement(9781..10938)
                     /gene="bll6280"
     CDS             complement(9781..10938)
                     /gene="bll6280"
                     /note="ORF_ID:bll6280
                     unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC51545.1"
                     /db_xref="GI:27354559"
                     /translation="MTVLIDVDSDEHHLLSRAPGNGARAMASPGNRAECLDIGLINNM
                     SDAALMSTERQLFDLLAAAAGRLCVRLHFYTMEATPRSEWGRDYVRRYYRGIDDLLNR
                     RLDGVIVTGAEPRAASLTDEPYWTTFAEIADWARENTVSSVFSCLAVHGAVLHLDGVT
                     RHKLPTKCFGVFAQTKTRHHPLMQDVPRTFRIPHARWNEVQEEDLSDCGYSVLTWSAE
                     AGVDCFVKQQKNSLFVHFQGHPEYETQSLLGEYRRDMGRFLRGENEVCPTIPRGYLNK
                     GAEEILIAFRQKALSDRQPELFADFPADQLTRDLSNVWHLPAQHIYRNWLQYMVSQPA

  Query Match             12.5%;  Score 157.4;  DB 1;  Length 302650;
  Best Local Similarity   49.4%;  Pred. No. 6.4e-17;
  Matches  556;  Conservative    0;  Mismatches  551;  Indels   18;  Gaps    5;

Qy         49 TTCAAGCCGTTCGCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCTGAGA 108
Db     294417 TTCAATCCGCTGTCGCCGGACTTCATCCGCAACCCCTATCCGCATTACGACCGGCTGCGC 294358

Qy        109 GAGGCAACCCCCATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCACCCGATACCAC 168
Db     294357 GCGATCGATCCGATCCATGTGACGCCGTTCGGCCAGTTC---GTCGCCAGTCGCCACGCC 294301

Qy        169 GACGTGTCGGCGGTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGAGTGGGAATCG 228
Db     294300 GATGTCAGCCTCGTGATGCGCGACAAGCGCTTCGGCAAGGATTTCGTCGAGCGGTCCAAG 294241

Qy        229 AGCGCGGAGTACTCGTCGGCCATTCCCGAGC---TCAGCGATATGAAGAAGTACGGATTG 285
Db     294240 CGCCGCTACAGCGAAAAGATCATGGACGAGCCGGTGTTCCGCAGCATGAGCCACTGGATG 294181

Qy        286 TTCGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACG 345
Db     294180 CTGCAGGCCGATCCGCCCGACCACACCCGCCTGCGCGGCCTCGTGGTGAAGGCCTTCACC 294121

Qy        346 TCACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGAT 405
Db     294120 GCGCGGCGCGTCGAGGACATGCGGCCGCGGATCCAGGAGATCGTCGATGAGGCCATCGAT 294061

Qy        406 GCTCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATG 465
Db     294060 GCCGTGATCGACCGCGGCCACATGGACCTGATCGAGGATTTCGCCTTCCGCCTGCCCGTC 294001

Qy        466 CGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTC 525
Db     294000 ACCATCATCTGCGACATGCTGGGAATCCCCGAGGATCATCGCGAGGTCTT------CTAC 293947

Qy        526 GGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACC 585
Db     293946 AAGAGCTCGCGCGACGGCGGGCGGCTGCTCGACCCCGTGCCGCTCACGCCGGAGGAGATC 293887

Qy        586 AAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAG 645
Db     293886 GCGAAGGGCAACGCCGGCAACATGATGGCGCAGATGTATTTCCAGCAGCTGTTCGAGCTC 293827

Qy        646 CGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGAC 705
Db     293826 CGGCGCCGCAATCC---CGCCGACGACCTCACCACCCAGCTGGTGCAGGCCGAGGAAGAC 293770

Qy        706 GGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGC 765
Db     293769 GGCAACAAGCTCACCAACGAGGAGCTGACCGCCAACATCATCCTGTTGTTCGGCGCCGGC 293710

Qy        766 ACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAG 825
Db     293709 CACGAGACCACCGTCAATCTGATCGGCAACGGCCTCTTGGCGCTGCATCGCAACCCGGAC 293650

Qy        826 GCGCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTC 885
Db     293649 CAGCTCGCGCTGCTCAAGGCCCGGCCGGAGCTGATGGTCAATGCGATCGAGGAGTTCCTG 293590

Qy        886 CGCTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTAC 945
Db     293589 CGCTACGATTCCTCGGTGCAGATGACCGGGCGCGTCACGCTCGAGGATATCGACGATCTC 293530

Qy        946 TGCGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGA 1005
Db     293529 GGCGGCAGGAAAATCCCCAAGGGCGAGACCGTGCTGTGCCTGCTCGGCTCGGCCAACCGC 293470

Qy       1006 GATGGGACTGTATTCTCC---AGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCG 1062
Db     293469 GATCCGGCGGTCTATCCCGATCGCCCCGACCGGCTCGACGTCACCCGTCCCAACGTCAAG 293410

Qy       1063 AGCCTCGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAG 1122
Db     293409 CCGCTGTCGTTCGGCGGCGGCATCCATTTCTGCCTGGGGGCCCAATTGGCGCGGATCGAG 293350

Qy       1123 GCGGAGATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGATG 1167
Db     293349 GCCGAGATCGCCATCGCCACGCTGTTGCGCCGGCTGCCGGATCTG 293305

RESULT 12
SERCP450A
LOCUS       SERCP450A               3078 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  Saccharopolyspora erythraea ORF 1 gene, partial cds; cytochrome
            P-450 gene, complete cds; ORF 2 gene, partial cds.
ACCESSION   M83110
VERSION     M83110.1  GI:152682
KEYWORDS    cytochrome P450.
SOURCE      Saccharopolyspora erythraea
  ORGANISM  Saccharopolyspora erythraea
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
REFERENCE   1  (bases 1 to 3078)
  AUTHORS   Andersen,J.F. and Hutchinson,C.R.
  TITLE     Characterization of Saccharopolyspora erythraea cytochrome P-450
            genes and enzymes, including 6-deoxyerythronolide B hydroxylase
  JOURNAL   J. Bacteriol. 174 (3), 725-735 (1992)
```

```
  MEDLINE   92121109
   PUBMED   1732208
COMMENT     Original source text: Saccharopolyspora erythraea (library: NRRL
            2338) DNA.
FEATURES             Location/Qualifiers
     source          1..3078
                     /organism="Saccharopolyspora erythraea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1836"
                     /tissue_lib="NRRL 2338"
     CDS             <1..859
                     /note="ORF homologous to Escherichia coli and Salmonella
                     typhimurium dnaE gene; putative"
                     /codon_start=2
                     /transl_table=11
                     /protein_id="AAA26482.1"
                     /db_xref="GI:152683"
                     /translation="AGGGRGHRAQAPAGARPVRPVRRRGRRRRRGVVVAAGAPAVHAG
                     RVAAQADALLRAGDAGPLRLGPPSGRGRAAARAVPGHRHRRARRRRARGGQGPGQDRR
                     DDLGDPAPHQQERPPVGDRDPGGPRRQRGGAVLPEVLRDVRRLPARGHRDRGEGPHQR
                     ARGHHQHLRLRRGAGGHLGGRDRPGNQPRVRDQGAGEQGGPVAGRRAQAHLAGALGHG
                     PGAREAAGAARGHPPRAVQRLLRVDRERAAGRAEGAARRGVLRDGAVSRRGAASPLRA
                     PRPLRLRGS"
     CDS             949..2166
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P-450"
                     /protein_id="AAA26483.1"
                     /db_xref="GI:152684"
                     /translation="MTTGEVPDLLAFDDAFAQDRHNRYARMREEPVQRIRTVNGLDAW
                     LITRYEDVKQALLDPRIAKDFGRTQQIIEKRLADAERRPGFSPDLGPHMLNTDPPDHT
                     RLRKLVVKAFTARRVEGLRPRIEQITDDLLDRLAGRSEVDLIDEFAFPLPITVISELM
                     GVEDSRRDDFRSWTNVLVDGSQPEAQAQASVAMVEYLTELIAKKRTEPGDDLLTALLE
                     AVEDGDRLSEGELIAMVFLLLVAGHETTVNLIGNCVLSLLGNPDQLAALRNDPSLLPG
                     AIEETLRYESPVANGTFRHTAEAVRFGDVVIPEGELVWVALGAANRDGERFEDPDRFD
                     ITRETTGHVAFGHGIHFCVGAALARLEAQIAVGRLLERFPDLRMAASPDDLRWRFSVL
                     MRGLEKLPVRPGA"
     CDS             complement(2365..>3078)
                     /note="ORF 2"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26484.1"
                     /db_xref="GI:152685"
                     /translation="LIAVATGSTLFVLSQYLQLVRGHSAVVAGMAALPLAAGSVVGSA
                     LGARAPARIGYRACIVTGFAVTAAGFGVLAALGPDSGQPHIAFGLLLCGFGTGFAGPA
                     ATSTALGAVPADRAGMGSALNDTHQQLGIAFGVAVLGGLLSTAYRAFLPTGVPHDAST
                     SLAATLSFADERTSAALADAARLAFTQAQSATMTAGLACALAGAAVAMLSLRSGRASA
                     VALTAVSATARRNLRAGGT"
BASE COUNT      463 a   1060 c   1137 g    418 t
ORIGIN

  Query Match             12.2%;  Score 154;  DB 1;  Length 3078;
  Best Local Similarity   49.1%;  Pred. No. 5.3e-16;
  Matches  569;  Conservative    0;  Mismatches  540;  Indels   51;  Gaps    4;

Qy         45 CGACTTCAAGCCGTTCGCGCCTGGGTACGCGGAGGACCCGTTCCCCGCGATCGAGCGCCT 104

Db        969 CGACTTACTGGCTTTCGACGACGCCTTCGCGCAGGACCGGCACAACCGGTACGCGCGGAT 1028

Qy        105 GAGAGAGGCAAC------CCCCATCTTCTACTGGGATGAAGGCCGCTCCTGGGTCCTCAC 158

Db       1029 GCGCGAGGAGCCGGTGCAGCGCATCCGCACGGTGAACGGCCTCGACGCGTGGCTGATCAC 1088

Qy        159 CCGATACCACGACGTGTCGGCGGTGTTCCGCGACGAACGCTTCGCGGTCAGTCGAGAAGA 218

Db       1089 CCGCTACGAGGACGTCAAGCAGGCGCTGCTGGACCCCAGGATCGCCAAGGACTTCGGGCG 1148

Qy        219 GTGGGAATCGAGCGCGGAGTACTCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTA 278

Db       1149 CACCCAGCAGATCATCGAGAAGAGGCTGGCCGACGCCGAGCGCAGGCCGGGGTTCTCGCC 1208

Qy        279 CGGATTGTTCGGGCTGCCG------------CCGGAGGATCACGCTCGGGTCCGCAAGCT 326

Db       1209 CGACCTCGGCCCGCACATGCTCAACACCGACCCGCCGGACCACACCCGGTTGCGCAAGCT 1268

Qy        327 CGTCAACCCGTCGTTTACGTCACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCAC 386

Db       1269 GGTGGTCAAGGCGTTCACCGCGCGGCGGGTCGAGGGCCTGCGGCCGCGGATCGAGCAGAT 1328

Qy        387 CGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTA 446

Db       1329 CACCGACGACCTGCTCGACCGCCTGGCGGGCCGGTCCGAAGTGGATCTGATCGACGAGTT 1388

Qy        447 CGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGA 506

Db       1389 CGCGTTCCCGCTGCCCATCACGGTGATCTCGGAGCTGATGGGCGTGGAGGACTCGCGGCG 1448

Qy        507 CGAGAAGTTCCGTCGCTTCGGCTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCC 566

Db       1449 CGACGACTTCCGG---------TCGTGGACCAACGTCCTGGTCGACGGTCGCAGCCGGA 1499

Qy        567 CCAGGTCGATGAGGAGACCAAGACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCT 626

Db       1500 GGCGCAGGCCCAGGCGTCGGTGGCGATGGTCGAGTACCTGACCGAGCTGATCGCCAAGAA 1559

Qy        627 CCATGACGTCCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCT 686

Db       1560 GCGCACCGAGCCCGGTGAC-----------------------GACCTGCTGACCGCGCT 1595

Qy        687 GCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGG 746

Db       1596 GCTGGAGGCGGTCGAGGACGGCGACCGCCTCAGCGAGGGCGAGCTGATCGCCATGGTGTT 1655

Qy        747 TGCGATTATCGCTGCTGGCACCGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAA 806

Db       1656 CCTGCTGCTGGTCGCCGGGCACGAGACGACGGTCAACCTGATCGGCAACTGCGTGCTGTC 1715

Qy        807 CCTGCTGCGGTCGCCCGAGGCGCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAA 866

Db       1716 GCTGCTGGGCAACCCCGACCAGCTCGCCGCGCTGCGCAACGACCCGTCGCTGCTGCCCGG 1775

Qy        867 CGCGCTCGATGAGGTGCTCCGCTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGC 926

Db       1776 CGCGATCGAGGAGACCCTGCGCTACGAGAGCCCGGTGGCGAACGGCACGTTCCGGCACAC 1835

Qy        927 CAGGCAGGACCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCT 986

Db       1836 CGCCGAGGCGGTCCGGTTCGGCGACGTGGTGATCCCCGAGGGAGAGCTCGTCTGGGTCGC 1895

Qy        987 GATCCCGAGCGCCCTGAGAGATGGGACTGTATTCTCCAGGCCAGACGTGTTTGATGTGCG 1046

Db       1896 GCTCGGGGCGGCGAACCGGGACGGTGAGCGCTTCGAGGACCCCGACCGCTTCGACATCAC 1955

Qy       1047 ACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTC 1106

Db       1956 CCGGGAGACCACCGGGCACGTCGCCTTCGGCCACGGCATCCACTTCTGCGTGGGCGCCGC 2015

Qy       1107 CCTTGCTCGCCTCGAGGCGGAGATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGAT 1166

Db       2016 CCTGGCCCGGCTGGAGGCCCAGATCGCCGTCGGCAGGCTCCTGGAGCGCTTCCCGGACCT 2075

Qy       1167 GAAGCTGAAAGAAACTCCCG 1186

Db       2076 GCGGATGGCCGCGTCGCCGG 2095


RESULT 13
AX697983
LOCUS       AX697983                1233 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 7 from Patent WO03010193.
ACCESSION   AX697983
VERSION     AX697983.1  GI:29499037
KEYWORDS    .
SOURCE      Micromonospora carbonacea
  ORGANISM  Micromonospora carbonacea
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
```

```
            Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE   1
  AUTHORS   Farnet,C.M., Staffa,A. and Yang,X.
  TITLE     Genes and proteins for the biosynthesis of rosaramicin
  JOURNAL   Patent: WO 03010193-A 7 06-FEB-2003;
            Ecopia Biosciences Inc. (CA)
FEATURES             Location/Qualifiers
     source          1..1233
                     /organism="Micromonospora carbonacea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:47853"
BASE COUNT      225 a    452 c    377 g    179 t
ORIGIN

  Query Match             11.9%;  Score 149.4;  DB 6;  Length 1233;
  Best Local Similarity   49.7%;  Pred. No. 3.6e-15;
  Matches  437;  Conservative    0;  Mismatches  436;  Indels    6;  Gaps    2;

Qy        288 CGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTC 347
Db        261 CGGCATCGACCCGCCGCGGCACGCGCCGCTGCGCAAGCTGGTGAGCCAGGCGTTCACCCC 320

Qy        348 ACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGC 407
Db        321 CCGCCGGATCGCCCAGATGGAGCTGCGGATCGGGCAGATCACCGCCGACGTGCTCGACCA 380

Qy        408 TCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCG 467
Db        381 GGTACGCGACCAGGACCGGATCGACATCGCCAGCGACCTCGCGTACCCGCTGCCGGTGAC 440

Qy        468 CGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGG 527
Db        441 GGTCATCGCCGAGCTGCTCGGCATTCCCACCAAGGATCACGAGAAGTTCCGCGAGTGGGT 500

Qy        528 CTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAA 587
Db        501 GGACATCATCCTCAGCAACGAAGGGCTGGAGTATCCCAACCTCCCGGACGACTTCACCGA 560

Qy        588 GACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCG 647
Db        561 GACGGTGGGCCCCGCCATCGAGGAGTGGTCCGAATTCCTGTACGCCCAGATCGCCCACAA 620

Qy        648 GCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGG 707
Db        621 GCGCGCCGAACCG---AAGGACGACCTGATCAGCGGCCTCTGTGCGGCGGAGGTCGACGG 677

Qy        708 CAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCAC 767
Db        678 GCGCAAGCTGACCGACGAGGAAGTCGTCAACATCGTCGCGCTGCTGCTCACCGCCGGGCA 737

Qy        768 CGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGC 827
Db        738 CATCTCCAGCGCCACGCTGCTCAGCAACCTGTTCCTGGTGCTGGAGGAGCACCCGCAGGC 797

Qy        828 GCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCG 887
Db        798 ACAGGCCGCGGTCCGCGCCGACCGCAGCCTCGTGCCGGGCGTGATCGAGGAGACGCTGCG 857

Qy        888 CTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTG 947
Db        858 CT---ACCGGTCCCCGTTCAACTGCATCTTCCGGATCCTGAACGAGGACACCGACATCCT 914

Qy        948 CGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGA 1007
Db        915 CGGCCACCCCATGCGCAAGGGCCAGATGGTGATCGCCTGGATCGCCTCCGCGAACCGCGA 974

Qy       1008 TGGGACTGTATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCT 1067
Db        975 CACCGAGGTGTTCACGGACCCGGACACCTTCGACATCCGACGCGAGTCGAACAAGCACCT 1034

Qy       1068 CGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGA 1127
Db       1035 GGCGTTCGGCCACGGCATCCACCACTGCCTGGGCGCGTTCCTGGCCAGGCTGGAGGCGAA 1094

Qy       1128 GATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGAT 1166
Db       1095 GGTCTTCCTCAACCAGACGCTCGACCAGTTCACCGAGTT 1133

RESULT 14
AX697977/c
LOCUS       AX697977               60196 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 1 from Patent WO03010193.
ACCESSION   AX697977
VERSION     AX697977.1  GI:29499034
KEYWORDS    .
SOURCE      Micromonospora carbonacea
  ORGANISM  Micromonospora carbonacea
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE   1
  AUTHORS   Farnet,C.M., Staffa,A. and Yang,X.
  TITLE     Genes and proteins for the biosynthesis of rosaramicin
  JOURNAL   Patent: WO 03010193-A 1 06-FEB-2003;
            Ecopia Biosciences Inc. (CA)
FEATURES             Location/Qualifiers
     source          1..60196
                     /organism="Micromonospora carbonacea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:47853"
BASE COUNT     6934 a  20042 c  23822 g   9398 t
ORIGIN

  Query Match             11.9%;  Score 149.4;  DB 6;  Length 60196;
  Best Local Similarity   49.7%;  Pred. No. 1.8e-15;
  Matches  437;  Conservative    0;  Mismatches  436;  Indels    6;  Gaps    2;

Qy        288 CGGGCTGCCGCCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTC 347
Db       3601 CGGCATCGACCCGCCGCGGCACGCGCCGCTGCGCAAGCTGGTGAGCCAGGCGTTCACCCC 3542

Qy        348 ACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGC 407
Db       3541 CCGCCGGATCGCCCAGATGGAGCTGCGGATCGGGCAGATCACCGCCGACGTGCTCGACCA 3482

Qy        408 TCGCTCCGGACAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCG 467
Db       3481 GGTACGCGACCAGGACCGGATCGACATCGCCAGCGACCTCGCGTACCCGCTGCCGGTGAC 3422

Qy        468 CGCGATCAGCGCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGG 527
Db       3421 GGTCATCGCCGAGCTGCTCGGCATTCCCACCAAGGATCACGAGAAGTTCCGCGAGTGGGT 3362

Qy        528 CTCGGCGACTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAA 587
Db       3361 GGACATCATCCTCAGCAACGAAGGGCTGGAGTATCCCAACCTCCCGGACGACTTCACCGA 3302

Qy        588 GACCCTGGTCGCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCG 647
Db       3301 GACGGTGGGCCCCGCCATCGAGGAGTGGTCCGAATTCCTGTACGCCCAGATCGCCCACAA 3242

Qy        648 GCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGG 707
Db       3241 GCGCGCCGAACCG---AAGGACGACCTGATCAGCGGCCTCTGTGCGGCGGAGGTCGACGG 3185

Qy        708 CAGCAGGCTGAGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCAC 767
Db       3184 GCGCAAGCTGACCGACGAGGAAGTCGTCAACATCGTCGCGCTGCTGCTCACCGCCGGGCA 3125

Qy        768 CGATACCACGATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGC 827
Db       3124 CATCTCCAGCGCCACGCTGCTCAGCAACCTGTTCCTGGTGCTGGAGGAGCACCCGCAGGC 3065

Qy        828 GCTCGAGCTGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCG 887
Db       3064 ACAGGCCGCGGTCCGCGCCGACCGCAGCCTCGTGCCGGGCGTGATCGAGGAGACGCTGCG 3005
```

```
Qy         888 CTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTG 947
                || ||     ||    |   | |   | || ||   |  ||||   ||   |
Db        3004 CT---ACCGGTCCCCGTTCAACTGCATCTTCCGGATCCTGAACGAGGACACCGACATCCT 2948

Qy         948 CGGGGCATCGATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGA 1007
               |||     ||   || ||  ||||||  |   | ||||  |    |||     | ||
Db        2947 CGGCCACCCCATGCGCAAGGGCCAGATGGTGATCGCCTGGATCGCCTCCGCGAACCGCGA 2888

Qy        1008 TGGGACTGTATTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCT 1067
                        || ||| |     || |||   || ||  | ||||| ||  ||  |   |  |||
Db        2887 CACCGAGGTGTTCACGGACCCGGACACCTTCGACATCCGACGCGAGTCGAACAAGCACCT 2828

Qy        1068 CGCGTACGGTAGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGA 1127
                     |||| |||     |||  |||    |||||  || | || ||| ||   | || |||||| |
Db        2827 GGCGTTCGGCCACGGCATCCACCACTGCCTGGGCGCGTTCCTGGCCAGGCTGGAGGCGAA 2768

Qy        1128 GATCGCCGTGGGCACCATCTTCCGTAGGTTCCCCGAGAT 1166
               | ||  |  |    |   |    ||     |||| ||||| |
Db        2767 GGTCTTCCTCAACCAGACGCTCGACCAGTTCACCGAGTT 2729


RESULT 15
SAU421825/c
LOCUS       SAU421825               66808 bp    DNA     linear   BCT 16-MAR-2002
DEFINITION  Stigmatella aurantiaca ORF8, ORF7, ORF6, ORF5, ORF4, ORF3, ORF2,
            ORF1, stiA gene, stiB gene, stiC gene, stiD gene, stiE gene, stiF
            gene, stiG gene, stiH gene, stiJ gene, stiK gene, stiL gene and
            ORF9.
ACCESSION   AJ421825
VERSION     AJ421825.1  GI:19572309
KEYWORDS    acyl-CoA binding protein; cellulase; cytochrome P450 monooxygenase;
            kinase; methyl transferase; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6;
            ORF7; ORF8; ORF9; polynucleotide adenyltransferase; potassium
            channel beta chain; ribosome binding factor; stiA gene; StiA
            protein; stiB gene; StiB protein; stiC gene; StiC protein; stiD
            gene; StiD protein; stiE gene; StiE protein; stiF gene; StiF
            protein; stiG gene; StiG protein; stiH gene; StiH protein; stiJ
            gene; StiJ protein; stiK gene; stiL gene.
SOURCE      Stigmatella aurantiaca
  ORGANISM  Stigmatella aurantiaca
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE   1
  AUTHORS   Gaitatzis,N., Silakowski,B., Kunze,B., Nordsiek,G., Blocker,H.,
            Hofle,G. and Muller,R.
  TITLE     The biosynthesis of the aromatic myxobacterial electron transport
            inhibitor stigmatellin is directed by a novel type of modular
            polyketide synthase
  JOURNAL   Online Publication
  REMARK    J. Biol. Chem., 10.1074/jbc.M111738200
REFERENCE   2  (bases 1 to 66808)
  AUTHORS   Muller,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124
            Braunschweig, GERMANY
FEATURES             Location/Qualifiers
     source          1. .66808
                     /organism="Stigmatella aurantiaca"
                     /mol_type="genomic DNA"
                     /strain="Sg a15"
                     /db_xref="taxon:41"
     CDS             complement(4978. .6231)
                     /note="ORF8"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P450 monooxygenase"
                     /protein_id="CAD19077.1"
                     /db_xref="GI:19572310"
                     /translation="MLRSVSASRPPSPANPVAAFDLARLDDAFYADPFPLYRAMRERD
                     PVHRMPDGSLFLTRWADLDCIYRDTRTFSSDKRAEFGAKYGDAPLFEHHTTSLVFNDP
                     PLHTRVRRLIVGALTPRALSTMEPGLRTLVDRLLDGLAVKGAADLIEDFAAAIPIEVI
                     GNLLDIPTEERGPLRGWSLAILGALEPRLTAEQEACGNEAVTEFLDYLRILVAQRRAR
                     PGDPATDVLTRLIQGESDGERLTETELLHQCVFLLNAGHETTTNLIGNALELLARFPD
                     ERARLLRAPALIPTAVEEVLRYESSNQLGNRRVAEDTEIGGVAVPTGTFLTLCIGAAN
                     RDPARFEDPEHFDVGRQPNRHLAFGAGHTCAGMNLARIEARIALAAFLARFPDYALT
                     APPVRARRARFRGFTAMPTRLGSLR"
     CDS             6369. .7730
                     /note="ORF7"
                     /codon_start=1
                     /transl_table=11
                     /product="protein kinase"
                     /protein_id="CAD19078.1"
                     /db_xref="GI:19572311"
                     /translation="MEELISEALPDGSRVGPWRIEGRAGYGTYGAVYRAHRVGRICPR
                     SVALKLARYPNDLRFEREAELLERVRHANVPRLLGHGTWRAGPQGEAHPYVVMQWVEG
                     VRLYQWNKERLRTAREVLRVLAQVARALEATHMRGGLHRDVKGENVLVSPEGHASLMD
                     FGCGTWEGAAPLTGGILAPGTKAYRSPEALRFHWNHRNEADAHYQATPADDVYALGVT
                     AYRLCTGIYPPPATEPSGVGEDERETQEVKKPPSQVRPLVPALEQLILRMLRDNPKER
                     GRAGELAAEMEALAERAGAEADEPMNPNHSSVPSEITGRPVSVQREWGGWPVALVLVA
                     VGLLLWGSGSLNTGRRVAPASRMTEGGTGGVADAAVEALPVSSAAREPEPSGLNLDML
                     KEPLPGQRRPPCPRFQISIQGGCWLEVPQSSPPCGESYYEWKGACYSPVAAPPRPRTS
                     DPQ"
     CDS             complement(7849. .8577)
                     /note="ORF6"
                     /codon_start=1
                     /transl_table=11
                     /product="methyl transferase"
                     /protein_id="CAD19079.1"
                     /db_xref="GI:19572312"
                     /translation="MSQNIYDDPSFFESYSQLARSLKGLAGAPEWPALRALLPDVKDL
                     RVVDLGCGFGWFCRWARKNDAAHVLGLDLSEKMLTRAKAETSGDAITYQQADLERLVL
                     PEGAFDLAFSSLAIHYVEELEPLLATVHRGLTPGGWFVFSIEHPIFMASRQPEWMTGP
                     EGHRTWPVDSYQLEGKRTTHWLAEEVVKYHRTLGTTLNALLEQGFTLRHIEEWGPTDA
                     QLSAHPELAEERDRPMFLLVSAQR"
     CDS             complement(8622. .8879)
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="acyl-CoA binding protein"
                     /protein_id="CAD19080.1"
                     /db_xref="GI:19572313"
                     /translation="MALEDDFRSAQERVKALKTRPSNDRLLELYSLFKQATEGDVQGK
                     RPGILDVKERAKYDAWVGRKGLGRESAMQQYVALVERLLRG"
     CDS             8961. .10070
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="potassium channel beta chain"
                     /protein_id="CAD19081.1"
                     /db_xref="GI:19572314"
                     /translation="MVYLLCRSRSQLRWILNGSAHRVRFAWGGHARPRRMHFRHLGRS
                     GLMVSEISYGNWLTHGSQVEEAAAVACVRAALDEGITTFDTADVYAGTRAEEVLGRAL
                     QGQRREGLEIFTKVYWPTGPGPNDRGLSRKHILESIHGSLRRLKTEYVDLYQAHRYDT
                     ETPLEETMQAFSDVVRQGKAHYIGVSEWTADQLREGVKLSRQMGFQLVSSQPQYSMLW
                     RVIEAEVVPASQELGVGQVVYSPIAQGVLTGKYRPGQKPPEGSRATDEKGGASFIKRF
                     LNDDVLTRVQQLTPVAQEAGLSLAQLAVAWVLQNPNVSSAIIGASRPEQVKENVKAAG
                     VKLSEEAMRHIDTLLGPIVERNPKKTLENSPRKRP"
     CDS             10428. .10862
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /product="ribosome binding factor"
                     /protein_id="CAD19082.1"
                     /db_xref="GI:19572315"
                     /translation="MTSFKQRRRRAPRRGTESSLSSFEAQSLPAHHLRVQSTLFQEVA
                     RLFRGELSDPLLEGVSVTSFELTPDGRNARIGFTLPPEYAEGGSAPIEAALVRASGFI
                     RAQLALGLNLKRVPHLRFVCVGVAPPGFSIPEDGSAEEGGES"
     CDS             10859. .12016
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD19083.1"
                     /db_xref="GI:19572316"
                     /translation="MRPRRIEGAPGAVPILAWARAVPAGAVKQLQLLARQPYLAEHVA
                     AMPDVHVAEGVAVGTVFATERHIVPGALGGDLGCGVSAVRFDFPAMSLGRVELTRLLA
```

```
                     AFARAVPVGDAVHRGRGLSLPPGLDIGRLSTHRLQREWERLVPKHLGTLGGGNHFLEL
                     DRDAEGALWLLLHTGSRGVGAAIAAHHLRVAQVSGEGSLPGLRTDTSEGAACLADTEL
                     ACQFARANRDTLLQRAVEVLAGAIGQGPVLDSRVDVHHNHVAAESHFGRTLLVHRKGA
                     VGLVAGQQGLIPGSMGTASYLVAGKGEPQAFCSCSHGAGRVMTRREARERIRPAALAH
                     ALRRVVFDEGRVTALVEEAPTAYRDITEVLEDEAELVTPLRRLTPIAVLKG"
     CDS             complement(12082. .13974)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="cellulase"
                     /protein_id="CAD19084.1"
                     /db_xref="GI:19572317"
                     /translation="MKNSLLLLMLAVLPLYGTAALARPGMANAEPGHFEGPENAGLTA
                     AQCNPRVPLSTRGRYIVDRCGERFKLKSVNWFGASDQLEVVGGLDKQKLSDLVAGMKA
                     MGFNSVRLPFSNNMLHPDDNKPAAQRCLEKYSVTCIDSTKNPELLNKTALEVYDAVVA
                     ELTRQGLVVILNNHTTKSMWCCGWDGNGFWDSSQALQRWQDDWVMLAARYQGNKWVAG
                     ADLRNEVRPDGLDSPNWGMRNQHDWHMAAQTMGNLLLRTNPDLLIVVEAVNWWGLLDG
                     SRPQLKPVRQRPVALLRGDKLVYAVHNYGYTGPNQSGGSLGSGPKYSDMDKPTLYGTL
                     DQEWGFALAANQAYTAPVWMSEFGVGYNEQAANSRAWFSNLADYLIEKDVDWAYWAIN
                     AAKLQNSVQGEDETYGLWSYPDWNGVRSGDWRLGTGPLGRLLGADGRTGAVDATRFLP
                     MTVLLKDGSSTYYVDNNSLDSDWHSGKAKISCPRGYRMVGVSANFTILCTNAGAVPAQ
                     QGTGSYQTITQEVSPLRYPGDWANQSIKFECPTGFYAVGISTANPFFLELAGLLCEQN
                     TGGIPLNSKSAKNFWGGDQRASSSGGDWSVSQYKGQCADNQYLIGLAHRRSGLADYPS
                     VVLCSN"
     gene            14345. .21466
                     /gene="stiA"
     CDS             14345. .21466
                     /gene="stiA"
                     /function="PKS"
                     /codon_start=1
                     /transl_table=11
                     /product="StiA protein"
                     /protein_id="CAD19085.1"
                     /db_xref="GI:19572318"
                     /translation="MTQLASRLGVGSESIDPRERFSRYGIDSRDTKSLLAELGALLKR
                     PLSPTLAWTYPTVEALSRHLAGEADVSGGERADTLGVDQDEPIAIIGMAGRFPKAPDV
                     ESFWSLLRDGVDAITEVPPERWDLDAFYSVGLKSPGKMNTRWGGFLDQVDQFDPLFFG
                     ISPREAIQMDPQQRLALELSWEALESAGVAVDSLMDSRTGVFMGAMWSDYAKLTRGAL
                     GLIGPHTATGQDSSIIAARVSYFLGLRGPSLTVNTACSSALVAVHLACQSLRNGEAAM
                     ALAGGVNLLLSPESSVAMSQLGALSPDGRCKAFDARANGYVRGEGGGVVVLTPLSRAI
                     ANGYPVLCVIRGSAVNNDGFSNGLTAPNPRAQEEVLRDAYARAGVSPDKVHYVETHGT
                     GTVLGDPIEANALSAVLGVNRPAERPLILGSVKTNIGHLEAASGIAGLMKVVLSIQHR
                     AIPPNLHFEQPNPYIRFEELRLLVPKSLHPWPNEGESALAGISSFGFGGTNCHLVLEA
                     PSPGSAHIAPLAAETPEELSLLARQARDAVSVVGSQVELADLCFSVAARCREGSQRLA
                     LIARSPAELAAQVDGFLAGEKRPGLLRGEVTSGGSQKPVFVFSGTGSQWPGMGRELLR
                     TEPVFRGAIERCDRALRGLSGWSMLEELSAGEASSRLKECSGLQTAIFAIQVGLAALW
                     RSWGIEPAGVVGHSIGEVAAAHVAGILNLEDAARVIHHLSRLMGRLSGQGAMIVVDLS
                     AEEANRLLREWERHVVVAACNAPASTVLSGAREVIEELGAELERRGIAAHTVRIDTPS
                     HSPLMDPFMGELAEVLAGLSPRVAAVPMVSTVTGAFLKAEECTAAYWVRNLREPVLFS
                     QALTKLFEQDLRLFLELSPHPILARSVEQCLKGEGRQGVVLASMRRGENEGACMRESL
                     GRLFTLGTSVAWQALYPAEAKRVELPMGLTGSLDAELRGSARLLPLSAHTAEALEDA
                     ARALQASIAKHPDVRLDDLCYSASVGKSVHAHRLAFVTDSKRALVEHLDDFLKGESRP
                     GRIAGVCATGEPKKVVFVFPGQGSQWQGMGRQLLRQEPVFREAIEECDAAIRSCAGWS
                     LIDALQADEPHSRLGSIEVIQPTLFAIEVALAALWRSWGVVPDAVVGHSMGETAAACV
                     AGALRLTDAARVICERSQLLRRVAGQGGMALVELSLEQAREMLSRKDGRVSIGASNSP
                     RSTVLSGDAIALEEILRELEAQGTFCRRVKVDVASHSPQVDPLKDDLLRVLSTIAPRA
                     ATVPLYSTVTGEPIEGARMDASYWMRNLREPVLFGQAIERLAGDGHGVFVELSPHPIL
                     LPSIDQTLRHKSHKGVVLPSLRREEPERAVLLESLGTLFTGGGRVDFRGLHPLGGHRV
                     RLPHLPWQRQRFWLDMDVAAAGAASVGAAWHPGGGSEHPLLGHRLPSLAHAPACHVWE
                     NKLSRSAVSYLDDHRVGGTVVLPGTSYIEMALAAAAKGSGKKTVSLRDVELREKMSFL
                     ETASRTVQLSLIQEEAGHSSFRLHGRSTPDEPWTLHGGGTVHVEEGDGPPSADPVLWA
                     AVETVRLQGKEVSASDAYEAMHAQGNEYGPLFRGIEQMWVRDAELVARLRVPEPLEQE
                     LPEYRFHPAALDLGLQALIVLAAEHGCEGLALPTRLESIRAHASPGARMWCHVRLNPV
                     REAASFSGEVLFLDESGQRLAELQGARFQILGEGSRPHDGGRLVDWLHEVEWQPAPVP
                     VPTAEPSSQGRWVLFADGTGVAEGLAAQIAARGEPCIRVLPGMAFAEAEDGHFVVDPE
                     QPEDVRKLFGALSRTSTIPLRGIVHLWSLDAPSPEEGGVDALSRARRLGSGSVLSLIQ
                     GLNAAELAQKPRLWLVTRGVQAVGSAGPGAGAVAQSPLWGLGQVVAYEHAELRCTRVD
                     LAPVSVPGEEQTLLQELLAPAGDDQVALRGRERHVPRLTRCRTDEVAGRSTDLRAGAI
                     RSDGCYLITGGLGGLGLTIAKWLVARGARHLVLVGRRGASDAAEIALGQMRGAGAECL
                     VMKADVANTSEVEELVGTIQRKMPPLRGIFHAAGLLDDGTVLQLTQARLRSVMAPKVE
                     GAWNLHRATLGASLDFFVLFSSAASLVGIPGQGNYAAGNAFLDALAGLRKARGLACQS
                     IQWGPWSEIGLAAEQDNRGERLALRGMGSLNPREGLEALEALLSPGEPLRERARIGVL
                     RLDARRWCKSYPAIRSSPLFSVLLAGERNAKQVLERSGLRARLNEEPARTRRAALESF
                     LREEAAQVLRLPPARIDLSRSLQTLGMDSLLSLELRNRLEANLELTLPATVLWNYPTL
                     SALAPFLAERMGIQLNGVEEPPQAVESPSVSEEDGAAMLKMLEQFQELSNEDLLGMLS
                     DDKEDEVSS"
     gene            21463. .26220
                     /gene="stiB"

  Query Match             11.8%;  Score 149.2;  DB 1;  Length 66808;
  Best Local Similarity   49.2%;  Pred. No. 1.9e-15;
  Matches  450;  Conservative    0;  Mismatches  458;  Indels    6;  Gaps    2;

Qy        298 CCGGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTCACGCGCCATC 357
Db       5928 CCGCCGCTGCACACGCGGGTGCGCCGACTGATCGTGGGCGCGCTGACGCCGCGCGTTG 5869

Qy        358 GACCTGCTGCGCGCCGAAATACAGCGCACCGTCGACCAGCTGCTCGATGCTCGCTCCGGA 417
Db       5868 TCGACAATGGAGCCAGGGTTGCGCACCCTGGTGGACCGGCTGCTGGACGGGCTGGCCGTG 5809

Qy        418 CAAGAGGAGTTCGACGTTGTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGC 477
Db       5808 AAGGGCGCCGCAGACCTGATCGAGGACTTCGCCGCCGCCATCCCCATCGAGGTAATTGGC 5749

Qy        478 GCTCTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGACT 537
Db       5748 AACCTGCTGGACATCCCGACGGAGGAGCGGGGGCCGCTGCGGGGCTGGTCGCTCGCGATC 5689

Qy        538 GCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACCAAGACCCTGGTC 597
Db       5688 CTCGGCGCGCTCGAACCCCGTCTGACGGCCGAGCAGGAGGCCTGCGGGAATGAGGCGGTA 5629

Qy        598 GCGTCCGTCACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGGCGCAGGAAC 657
Db       5628 ACAGAGTTCCTGGACTACCTCCGCATCCTCGTCGCGCAGCGGCGAGCGCGGCCGGGGGAC 5569

Qy        658 CCGCTCGAAAATGACGTCTTGACGATGCTGCTTCAGGCCGAGGCCGACGGCAGCAGGCTG 717
Db       5568 CC---CGCAACGGACGTGCTGACGCGCCTCATCCAGGGCGAATCCGACGGGGAACGATTG 5512

Qy        718 AGCACGAAGGAGCTGGTCGCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATACCACG 777
Db       5511 ACCGAGACGGAGCTGCTGCACCAGTGCGTCTTCCTCCTGAACGCGGGGCACGAGACGACA 5452

Qy        778 ATCTACCTTATCGCGTTCGCTGTGCTCAACCTGCTGCGGTCGCCCGAGGCGCTCGAGCTG 837
Db       5451 ACGAACCTCATCGGCAACGCCCTGGAACTGCTGGCGCGCTTCCCGGACGAGCGCGCTCGG 5392

Qy        838 GTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACAAT 897
Db       5391 CTGCTGCGAGCCCCCGCGCTGATCCCGACGGCAGTGGAGGAGGTCCTGCGTTACGAATCC 5332

Qy        898 ATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGACCTGGAGTACTGCGGGGCATCG 957
Db       5331 TCCAACCAGCTCGGCAACCGGCGGGTGGC---GGAGGACACGGAGATTGGCGGCGTAGCG 5275

Qy        958 ATCAAGAAAGGGGAGATGGTCTTTCTCCTGATCCCGAGCGCCCTGAGAGATGGGACTGTA 1017
Db       5274 GTGCCGACCGGCACCTTCCTCACGCTCTGCATCGGCGCCGCCAACCGCGACCCCGCGCGC 5215

Qy       1018 TTCTCCAGGCCAGACGTGTTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGT 1077
Db       5214 TTCGAGGACCCAGAGCACTTCGACGTGGGGCGCCAGCCCAACCGGCACCTCGCCTTCGCG 5155

Qy       1078 AGAGGCCCCCATGTCTGCCCCGGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCGTG 1137
Db       5154 GGCGGGGCGCATACGTGCGCGGGGATGAACCTTGCGCGGATAGAGGCGCGGATAGCGTTG 5095

Qy       1138 GGCACCATCTTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAAACTCCCGTGTTTGGATAC 1197
Db       5094 GCGGCGTTCCTCGCGCGCTTCCCGGACTATGCGCTGACCGCACCGCCGGTGAGGGCGCGG 5035

Qy       1198 CACCCCGCGTTCCG 1211
Db       5034 CGGGCGCGGTTCCG 5021
```

Search completed: October 4, 2003, 03:31:07
Job time : 4517.41 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         October  3, 2003, 11:40:01 ; Search time 8424.03 Seconds
                                            (without alignments)
                                            12307.996 Million cell updates/sec

Title:          US-10-014-717-1_COPY_7610_11875
Perfect score:  4266
Sequence:       1 gtggcggatcgtcccatcga..........gggaaatcattgccctatga 4266

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters:      45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  em_estba:*
                2:  em_esthum:*
                3:  em_estin:*
                4:  em_estmu:*
                5:  em_estov:*
                6:  em_estpl:*
                7:  em_estro:*
                8:  em_htc:*
                9:  gb_est1:*
                10:  gb_est2:*
                11:  gb_htc:*
                12:  gb_est3:*
                13:  gb_est4:*
                14:  gb_est5:*
                15:  em_estfun:*
                16:  em_estom:*
                17:  em_gss_hum:*
                18:  em_gss_inv:*
                19:  em_gss_pln:*
                20:  em_gss_vrt:*
                21:  em_gss_fun:*
                22:  em_gss_mam:*
                23:  em_gss_mus:*
                24:  em_gss_pro:*
                25:  em_gss_rod:*
                26:  em_gss_phg:*
                27:  em_gss_vrl:*
                28:  gb_gss1:*
                29:  gb_gss2:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                   %
Result            Query
  No.    Score  Match Length DB  ID                          Description
--------------------------------------------------------------------------
c  1      211    4.9    871  29  BZ566693                    BZ566693 pacs2-164
   2      188    4.4    753  29  BZ561977                    BZ561977 pacs2-164
c  3    180.2    4.2    979  29  BZ568491                    BZ568491 pacs2-164
   4      165    3.9   1303  29  BZ577421                    BZ577421 msh2_5402
   5    129.4    3.0    865  29  BZ566683                    BZ566683 pacs2-164
c  6    128.2    3.0    679  29  BZ562208                    BZ562208 pacs2-164
c  7      126    3.0   1348  29  BZ569008                    BZ569008 pacs2-164
   8    123.4    2.9    988  13  BX384927                    BX384927 BX384927
   9      123    2.9   4176  11  BC021544                    BC021544 Homo sapi
  10      123    2.9   4185  11  BC014631                    BC014631 Homo sapi
  11    118.8    2.8   1061  13  BQ069153                    BQ069153 AGENCOURT
  12    118.2    2.8   1201  13  BX332425                    BX332425 BX332425
  13    114.6    2.7   1798  29  AG171124                    AG171124 Pan trogl
  14    114.4    2.7    599  29  BZ560333                    BZ560333 pacs2-164
  15    113.4    2.7    899  13  BU528700                    BU528700 AGENCOURT
c 16    112.8    2.6   1798  29  AG171124                    AG171124 Pan trogl
  17      110    2.6    876  29  BZ573950                    BZ573950 msh2_3439
c 18      108    2.5   1101  29  AG039543                    AG039543 Pan trogl
  19    107.8    2.5   1196  13  BQ069593                    BQ069593 AGENCOURT
c 20    107.8    2.5   1198  29  CC190008                    CC190008 CH261-138
  21      107    2.5   1438  29  BZ557931                    BZ557931 pacs1-60_
c 22    106.6    2.5   1785  29  CC219595                    CC219595 CH261-62D
c 23    105.6    2.5   1281  12  BG852363                    BG852363 1024034A0
c 24    105.6    2.5   1695  29  CC290874                    CC290874 CH261-172
  25      105    2.5    499  12  BM626671                    BM626671 170006874
  26      105    2.5    577  12  BM646821                    BM646821 170006873
  27    104.6    2.5   1299  29  AG039481                    AG039481 Pan trogl
c 28    104.4    2.4   1610  29  BZ569386                    BZ569386 pacs2-164
  29    104.2    2.4    690  10  BB865539                    BB865539 BB865539
  30      104    2.4    692  12  BM652119                    BM652119 170006873
  31    103.6    2.4   1189  29  AG032118                    AG032118 Pan trogl
  32    103.4    2.4    968  13  BU557094                    BU557094 AGENCOURT
  33    102.8    2.4   1090  29  BZ562999                    BZ562999 pacs2-164
c 34    102.6    2.4    625  12  BM091001                    BM091001 ig21a10.y
c 35    102.2    2.4    766  10  BE299394                    BE299394 601118872
  36    102.2    2.4   1651  12  BG809816                    BG809816 mgct001xk
  37      102    2.4    600  29  BZ560280                    BZ560280 pacs2-164
  38      102    2.4    928  13  BQ957144                    BQ957144 AGENCOURT
  39    101.6    2.4    699  10  BF250997                    BF250997 EST418254
c 40    101.4    2.4   1542  29  AG032943                    AG032943 Pan trogl
  41      101    2.4    598  29  BZ560289                    BZ560289 pacs2-164
  42      100    2.3    704  28  AQ449661                    AQ449661 500003A11
  43     99.8    2.3    731  10  BG698437                    BG698437 602658472
c 44     99.8    2.3   1242  12  BM911414                    BM911414 AGENCOURT
c 45     99.8    2.3   1350  12  BM450237                    BM450237 AGENCOURT

                              ALIGNMENTS

RESULT 1
BZ566693/c
LOCUS       BZ566693                 871 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_6525.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_6525, genomic survey sequence.
ACCESSION   BZ566693
VERSION     BZ566693.1  GI:27196732
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 871)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol., (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
```

```
     source          1..871
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_6525"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      165 a    294 c    269 g    140 t      3 others
ORIGIN

  Query Match             4.9%;  Score 211;  DB 29;  Length 871;
  Best Local Similarity   57.6%;  Pred. No. 2.5e-27;
  Matches  396;  Conservative    0;  Mismatches  290;  Indels    1;  Gaps    1;

Qy        221 CTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAG 280
Db        768 CCTTCCTTGATGAGGTGGAAGGCTTCGACG-CGCGTTCTTCGGGATATCCGCTTCCGAGG 710

Qy        281 CGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGA 340
Db        709 CTCGCGAAATGAACCCCCCACAGCGGCTGCTGCTGGAGGTCGTCTGGGAATGCATGGAAC 650

Qy        341 ACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCG 400
Db        649 GGGCGGGGCTCTGCTCTGAACGCCTGAAAGGTTCCCGCACTGCGGTCTACGTCGGCATGC 590

Qy        401 GCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTC 460
Db        589 TGGGCATGGATTACCTTACCCTGCATGCGCGCGAGGCCGGTATCCCGGCGATCAATCCAT 530

Qy        461 ATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGC 520
Db        529 TCTATGCGGCTGGCAAGGAGTTCAGCTTCGCCGCCGGGCGCATCGCCTATCACCTTGGCG 470

Qy        521 TGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATC 580
Db        469 TGCACGGTCCGGCGATGGCGGTGAGCACCGCCTGCTCCTCTTCGCTGGTGGCGATTCACC 410

Qy        581 TGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGC 640
Db        409 TGGGTTGTCGCGCCTTGCAGAATCGCGAGGCGGACCTGGCGATAGCCGCTGGCGTCAATC 350

Qy        641 TGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACG 700
Db        349 TGATGCTGTCGCCCGAGCTGACTATCTACATGAGCCAGATCAAGGCGATTTCGCCGAGCG 290

Qy        701 GTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCG 760
Db        289 GACGCTGCCGGGTCTTCGACGCCGCCGCCGATGGCATCGTCCGCGGCGAAGGGGGTGGCG 230

Qy        761 TCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGA 820
Db        229 CGTTGCTGCTCAAGCGTCTGGAGGACGCTCGAGCCGACGGCGACCCGATCCTGGCGGTGA 170

Qy        821 TTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGA 880
Db        169 TTCGTGGCTCGGCGATCAACCAGGACGGCGAGAGCGCCGGCCAGACCGTGCCCAACGCCA 110

Qy        881 GCTCCCAAGAAATCGTGCTGAAACGGG 907
Db        109 ATGCCCAGGCCGCTGTGGGGCTGCAGG 83


RESULT 2
BZ561977
LOCUS       BZ561977                 753 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_3530.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_3530, genomic survey sequence.
ACCESSION   BZ561977
VERSION     BZ561977.1  GI:27183081
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 753)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol., (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..753
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_3530"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      127 a    250 c    243 g    133 t
ORIGIN

  Query Match             4.4%;  Score 188;  DB 29;  Length 753;
  Best Local Similarity   60.3%;  Pred. No. 2.7e-23;
  Matches  311;  Conservative    0;  Mismatches  205;  Indels    0;  Gaps    0;

Qy        633 GGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGGCGCTGGC 692
Db         95 GCTTCCCCTGATCCTTTCGCCCTATGGTTTCGTCCTGGTCAGCAAGTTGCGGGCGGTCGC 154

Qy        693 CAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAGGCGAAGG 752
Db        155 CAGCGATGGGCGTTGCAAGACCTTCGACGCCGCCGCCGACGGCTATGGCCGGGGAGAGGG 214

Qy        753 GTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATCGGATATT 812
Db        215 CTGCGTGGTGCTGGTTCTCAAGCGCCTGGCCGACGCCGAGCGTGACCGGGATCCGGTGCT 274

Qy        813 GGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTGACCGTGCC 872
Db        275 CGGGGTGATCCAGGGTAGCGCGCTCAACAACGACGGCTCCAGCAGCGGAATCACGGTTCC 334

Qy        873 GAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGCTGCGCCGC 932
Db        335 GAACATCCACGCCCAGCAAGCGGTGATCGCCGCTGCCTTGCGCAATGCCGGAGCCCAGGC 394

Qy        933 GTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGACCCCATCGA 992
Db        395 GAGCCAGGTCGACTACATCGAGGCACATGGTACCGGCACTCCGCTGGGCGACCCGATCGA 454

Qy        993 AATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCGCTGCTGAT 1052
Db        455 ACTGCGCGCGCTGGATGCCGTGCTCGGCGCGCAGCGCCCGGCTGACAGGCCGCTGCTGGT 514

Qy       1053 CGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACTGGGCTGCT 1112
Db        515 GGGATCGATAAAGGCCAACATCGGCCACCTCGAGCCAGTGGCCGGCGCTGCGGGACTGGC 574

Qy       1113 GAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCC 1148
Db        575 GAAGATCCTGCTCGCGTTGCAACACGACCAGTTGCC 610


RESULT 3
```

```
BZ568491/c
LOCUS       BZ568491                 979 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_7558.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_7558, genomic survey sequence.
ACCESSION   BZ568491
VERSION     BZ568491.1  GI:27201527
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 979)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .979
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_7558"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      187 a    324 c    299 g    164 t      5 others
ORIGIN

  Query Match             4.2%;  Score 180.2;  DB 29;  Length 979;
  Best Local Similarity   55.1%;  Pred. No. 6.5e-22;
  Matches  372;  Conservative    0;  Mismatches  297;  Indels    6;  Gaps    1;

Qy        186 CGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTT 245

Db        753 CCAAACCCGGTCAACCACCCCGGGGCCATGCAAGCCTTGGTCAGGCGCGTGAAAGGGGTT 694

Qy        246 CGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCG 305

Db        693 CACCNAGNACTTTTTCCGGATCTCGCCACGGNAAGCGGCAAGCATGGATCGGCAGCAACG 634

Qy        306 ACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCT 365

Db        633 CCTGTTCCTGATGGTTGCCTGGNGGCCCTGGAAGCGGGCAGGTATTCCGCCGCATGACTT 574

Qy        366 CGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCT 425

Db        573 GCGCGAACGCCAGGTCGGTGTTTTCGTCGGCGCCAATGCCCACGACTACGAGGCACGTTT 514

Qy        426 GCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAG 485

Db        513 GCTGG------CCTGCGCCGCCGGCATCGACGCGAACTACGGTACCGGCAGCGCGTTCTC 460

Qy        486 CGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGA 545

Db        459 GGCGATTTCCGGCCGACTGTCGCATTTCCTTGGCCTGCGCGGTCCCTCCCTGACCGTGGA 400

Qy        546 TACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGG 605

Db        399 CACGGCGTGCTCGTCATCGCTGACGGCTGTGCACCTGGCCTGCAACAGCCTGCTAGCCGG 340

Qy        606 GGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGT 665

Db        339 GGAGTGCGAGATTGCTATCGTCGGCGGGGTGAACGTCATCGCATCCTCGGCAATCTTCCA 280

Qy        666 GTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGA 725

Db        279 ATCCATGGGCGATGCCGGTGCGCTATCGGCGGACGGCACCTGCAAGACCTTCGACGATCG 220

Qy        726 GGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGG 785

Db        219 TGCCGATGGCTACGGCCGCGGCGAGGGTTGCGGCGTGGTGATCCTCAAGCGGGCGGAACG 160

Qy        786 AGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGA 845

Db        159 GGCGCGGGCGGACGGCGACCGGGTCATCGCCGTTCTGCTGGGCTCGGCGGTCAACCATGA 100

Qy        846 CGGTGCGAGCAGCGG 860

Db         99 CGGCGCCTGCGCGGG 85


RESULT 4
BZ577421
LOCUS       BZ577421                1303 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  msh2_5402.x1 msh Pseudomonas aeruginosa genomic clone msh2_5402,
            genomic survey sequence.
ACCESSION   BZ577421
VERSION     BZ577421.1. GI:27212482
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1303)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1303
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_5402"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."
BASE COUNT      278 a    362 c    359 g    295 t      9 others
ORIGIN

  Query Match             3.9%;  Score 165;  DB 29;  Length 1303;
  Best Local Similarity   57.9%;  Pred. No. 3.1e-19;
  Matches  307;  Conservative    0;  Mismatches  220;  Indels    3;  Gaps    1;

Qy        285 GCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGC 344

Db         94 GGGCATGGACCCGCAACAGCGAATCGCGCTCGAAGTTGCCTGGGAAGCGCTGGAGCATGC 153

Qy        345 CGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCC 404

Db        154 GGCTATCGACCCGTTTTCCCTCGATGGCAGCAAGCTGTGCGTGTTCATGGGGGCCAGTAC 213

Qy        405 GTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGG 464

Db        214 CAACGACTATGTACGGCTGCGCCAGCAACTCGGCGAGGTGGCTGACGTCAATGCGCA--- 270
```

```
Qy        465 CGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCG 524
Db        271 CCAGTTCTACGGCGAGACCAGCTTTATCGCCGGACGTATCGCCTACACCCTGNGCGGGCA 330

Qy        525 AGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGC 584
Db        331 AGGCCCGGCGATGATGATCGACACCTCCTGCTCGTCTTCGTTGGTGGCGCTGCACCAGGC 390

Qy        585 CTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGAT 644
Db        391 CTGCCAGAGCCTGCGCAGCGGCGAAAGCGACCTGNCGCTGGCCGGCGGNGTGAACCTGAT 450

Qy        645 GTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCG 704
Db        451 CCTTTCGCCCTATGGTTTCGTCCTGGTCAGCAAGTTGCGGGCGGTCGACAGCGATGGGCG 510

Qy        705 CTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGT 764
Db        511 TTGCAAGACCTTCGACGCCGCCGNCGACGGCTATGGCCTGNGAGAGGGCTGCGTTGTGCT 570

Qy        765 GGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATCGGATATTGG 814
Db        571 TGTTCTCAAGCGCCTGGCCGACGCCGAGCGTGACCGTGATCCGTGCTCGG 620


RESULT 5
BZ566683
LOCUS       BZ566683                 865 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_6520.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_6520, genomic survey sequence.
ACCESSION   BZ566683
VERSION     BZ566683.1  GI:27196700
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 865)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..865
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_6520"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      142 a     263 c     298 g     161 t       1 others
ORIGIN

  Query Match             3.0%;  Score 129.4;  DB 29;  Length 865;
  Best Local Similarity   53.5%;  Pred. No. 5.2e-13;
  Matches  296;  Conservative    0;  Mismatches  251;  Indels    6;  Gaps    1;

Qy        756 CGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATCGGATATTGGC 815
Db         45 CGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCGACCGGGTCATCGC 104

Qy        816 GGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAA 875
Db        105 CGTTCTGCTGGGCTCGGCGGTCAACCATGACGGCGCCTGCGCGGGGCTCACGGTACCCAA 164

Qy        876 CGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTC 935
Db        165 CGGACCGGCCCAGGAGGCGTTGATTCGCGAAGCGCTGGCCAACGCCAGCCTGTCGGCCGC 224

Qy        936 TTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGACCCCATCGAAAT 995
Db        225 GGCGGTGGGCTATGTCGAGGCCCATGGGACCGGCACAGTGCTCGGCGATCCGATCGAGTT 284

Qy        996 CCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGG 1055
Db        285 GCAGGCCCT------GGCCAACGCCTATCGAGGTACTGAGCGCGTGCCCTTGGCGGTGGC 338

Qy       1056 GTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAA 1115
Db        339 ATCGGTGAAAGGTAACGTCGGACATCTGGAGGCCGCCGCCGGTATCGCTTCGTTGATCAA 398

Qy       1116 GGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAA 1175
Db        399 GGCCTGCCTGGTGGTCGAGCGTGGGGTGATTCCCGGACAGGCTGGCTTCCAGCGTCCTAA 458

Qy       1176 CCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCGGACACCGTGGCC 1235
Db        459 CAGCCGCTTCGACTGGCAGGCAACGAGTCTCCAGGTGCCGCAGGCCAATCAGCCCTGGGA 518

Qy       1236 GGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGC 1295
Db        519 GGTGCCAGCGGCAGAACGGATCGCGGGCATCAGCGCTTTCGGTTTCACCGGTACCAATGT 578

Qy       1296 GCACGTGGTGCTG 1308
Db        579 CCATGTGCTGTTG 591


RESULT 6
BZ562208/c
LOCUS       BZ562208                 679 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_3644.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_3644, genomic survey sequence.
ACCESSION   BZ562208
VERSION     BZ562208.1  GI:27183746
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 679)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..679
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_3644"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      135 a     237 c     202 g     101 t       4 others
```

```
ORIGIN

  Query Match             3.0%;  Score 128.2;  DB 29;  Length 679;
  Best Local Similarity   59.3%;  Pred. No. 8.3e-13;
  Matches  239;  Conservative    0;  Mismatches  158;  Indels    6;  Gaps    1;

Qy       1623 CGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGA 1682
Db        515 CGCCTGGCTGTTTCCCGGCCAGGGCTGCCAGTATCCGGGAATGGCCAGGACGCTGTACGA 456

Qy       1683 TGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCT 1742
Db        455 CAGCGAGCTGCCGTTCCGTGAGGCATTGGACGAGTGTCTCGTGGCGTTGCGCGGGATGGG 396

Qy       1743 CGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCT 1802
Db        395 TGTCGAGCTGTCCTGCGCCTTGTT------CGAGTCGACGCCGGAGGCAGAGGCGTTGCT 342

Qy       1803 CGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCCGCGCT 1862
Db        341 GGCGCAGACACGCTATGCGCAGCCGGCGATTTTCGCCGTGGAATACGCGCTGGCCCGCTT 282

Qy       1863 GTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGT 1922
Db        281 GTGGATGGCCTGGGGCGTGCTCCCCGGCGCGCTGTTCGGTCACAGTTTCGGTGAAATTGG 222

Qy       1923 GGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCG 1982
Db        221 CGCCGCCTGTATCGCCGGGGCCCTCACGCTGGAGGACGCCCTGAGGATGGTCGAGGCGCG 162

Qy       1983 CGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTG 2025
Db        161 CGGACGCCTGGCCCAGGAACTGATGACCGCCGGAGGGGTGATG 119


RESULT 7
BZ569008/c
LOCUS       BZ569008                1348 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_8195.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_8195, genomic survey sequence.
ACCESSION   BZ569008
VERSION     BZ569008.1  GI:27202907
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1348)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..1348
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_8195"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      319 a    389 c    348 g    274 t     18 others
ORIGIN

  Query Match             3.0%;  Score 126;  DB 29;  Length 1348;
  Best Local Similarity   51.6%;  Pred. No. 2.1e-12;
  Matches  299;  Conservative    0;  Mismatches  279;  Indels    1;  Gaps    1;

Qy        129 CGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGA 188
Db        674 CGTCNGGCNGGTCNCCGCGGGCCGCTGGAATGGCG-AGCGATTTCACGACGCTGGTGCGA 616

Qy        189 TGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGA 248
Db        615 GCNGTCNCGGTACCACTATGCNNNGGAGNGGGCCTTCCTTGATGAGGTGGACGCTTTCGA 556

Qy        249 CGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACT 308
Db        555 CGCGCGTTTCTTCGGGATATCCGCTTCTGAGGCTCGCGAAATGGACCTCCAGCAGCGGCT 496

Qy        309 CTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGT 368
Db        495 GCTGCTGGAGGTCGTCTGGGAATGCATGGAACGGGCGGGGCTCTGCTCTGAACGCCTGAA 436

Qy        369 CGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCC 428
Db        435 AGGTTCCCGCACTGCGGTCTACGTCGGCATGCTGGGCATGGATTACCTTACCCTGCATGC 376

Qy        429 GCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGT 488
Db        375 GCGCGAGGCCGGTATCCCGGCGATCAATCCATTCTATGCGGCTGGCAAGGAGTTCAGCTT 316

Qy        489 CGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATAC 548
Db        315 CGCCGCCGGGCGCATCGCCTATCACCTTGGCGTGCACGGTCCGGCGATGGCGGTGAGCAC 256

Qy        549 GGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGA 608
Db        255 CGCCTGCTCCTCTTCGCTGGTGGCGATTCACCTGGGTTGTCGCGCCTTGCAGAATCGCGA 196

Qy        609 ATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTG 668
Db        195 GGCGGACCTGGCGATAGCCGCTGGCGTCAATCTGATGCTGTCGCCCGAGCTGACTATCTA 136

Qy        669 GCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTG 707
Db        135 CATGAGCCAGATCAAGGCGATTTCGCCGAGCGGACGCTG 97


RESULT 8
BX384927
LOCUS       BX384927                 988 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX384927 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
            Homo sapiens cDNA clone CS0DL004YA22 5-PRIME, mRNA sequence.
ACCESSION   BX384927
VERSION     BX384927.1  GI:30459273
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 988)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 8574.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DL004BA11QP1&cluster=8574.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
```

```
          Faraday Avenue Genoscope sequence ID : CS0DL004BA11QP1.
FEATURES             Location/Qualifiers
     source          1..988
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DL004YA22"
                     /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
                     25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      184 a    334 c    304 g    166 t
ORIGIN

  Query Match             2.9%;  Score 123.4;  DB 13;  Length 988;
  Best Local Similarity   53.0%;  Pred. No. 5.9e-12;
  Matches  290;  Conservative    0;  Mismatches  251;  Indels    6;  Gaps    1;

Qy        226 CTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTG 285

Db        328 CTGAAGGACCTGTCTAGGTTTGATGCCTCCTTCTTCGGAGTCCACCCCAAGCAGGCACAC 387

Qy        286 CGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCC 345

Db        388 ACGATGGACCCTCAGCTGCGGCTGCTGCTGGAAGTCACCTATGAAGCCATCGTGGACGGA 447

Qy        346 GCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCG 405

Db        448 GGCATCAACCCAGATTCACTCCGAGGAACACACACTGG------CGTCTGGGTGGGCGTG 501

Qy        406 TCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGC 465

Db        502 AGCGGCTCTGAGACCTCGGAGGCCCTGAGCCGAGACCCCGAGACACTCGTGGCTACAGC 561

Qy        466 GGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGA 525

Db        562 ATGGTGGGCTGCCAGCGAGCGATGATGGCCAACCGGCTCTCCTTCTTCTTCGACTTCAGA 621

Qy        526 GGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCC 585

Db        622 GGGCCCAGCATCGCACTGGACACAGCCTGCTCCTCCAGCCTGATGGCCCTGCAGAACGCC 681

Qy        586 TGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATG 645

Db        682 TACCAGGCCATCCACAGCGGGCAGTGCCCTGCCGCCATCGTGGGGGGCATCAATGTCCTG 741

Qy        646 TTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGC 705

Db        742 CTGAAGCCCAACACCTCCGTGCAGTTCTTGAGGCTGGGGATGCTCAGCCCCGAGGGCACC 801

Qy        706 TGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTG 765

Db        802 TGCAAGCCCTTCGACACAGCGGGGAATGGGTACTGCCGCTCGGAGGGTGTGGTGGCCGTC 861

Qy        766 GTCCTCA 772

Db        862 CTGCTGA 868


RESULT 9
BC021544
LOCUS       BC021544                4176 bp    mRNA    linear   HTC 20-MAY-2002
DEFINITION  Homo sapiens, clone IMAGE:3355480, mRNA.
ACCESSION   BC021544
VERSION     BC021544.1  GI:20987241
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4176)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 22 Row: p Column: 6
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1..4176
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3355480"
                     /tissue_type="Eye, retinoblastoma"
                     /clone_lib="NIH_MGC_16"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
BASE COUNT      702 a   1419 c   1319 g    736 t
ORIGIN

  Query Match             2.9%;  Score 123;  DB 11;  Length 4176;
  Best Local Similarity   50.2%;  Pred. No. 7.8e-12;
  Matches  340;  Conservative    0;  Mismatches  325;  Indels   12;  Gaps    1;

Qy       3128 TTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCGCTCTCGG 3187

Db       1584 TGGCCACCTCTGTCCTGCTGTCACCGGACTTCCTCTGGGATGTGCCTTCCAACTGGACGC 1643

Qy       3188 CGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTCGACAGAA 3247

Db       1644 TGGAGGAGGCGGCCTCGGTGCCTGTCGTCTACAGCACGGCCTACTACGCGCTGGTGGTGC 1703

Qy       3248 TAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGGGTCGGTC 3307

Db       1704 GTGGGCGGGTGCGCCCCGGGGAGACGCTGCTCATCCACTCGGGCTCGGGCGGCGTGGGCC 1763

Qy       3308 TCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCCGGCACGC 3367

Db       1764 AGGCCGCCATCGCCATCGCCCTCAGTCTGGGCTGCCGCGTCTTCACCACCGTGGGGTCGG 1823

Qy       3368 CCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAG-----------CG 3415

Db       1824 CTGAGAAGCGGGCGTACCTCCAGGCCAGGTTCCCCCAGCTCGACAGCACCAGCTTCGCCA 1883

Qy       3416 ATTCCCGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAG 3475

Db       1884 ACTCCCGGGACACATCCTTCGAGCAGCATGTGCTGTGGCACACGGGCGGGAAGGGCGTTG 1943
```

```
Qy       3476 ACGTCGTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGAT 3535
Db       1944 ACCTGGTCTTGAACTCCTTGGCGGAAGAGAAGCTGCAGGCCAGCGTGAGGTGCTTGGCTA 2003

Qy       3536 CGCACGGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGC 3595
Db       2004 CGCACGGTCGCTTCCTGGAAATTGGCAAATTCGACCTTTCTCAGAACCACCCGCTCGGCA 2063

Qy       3596 TGCGGCCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCG 3655
Db       2064 TGGCTATCTTCCTGAAGAACGTGACATTCCACGGGGTCCTACTGGATGCGTTCTTCAACG 2123

Qy       3656 AGCGGCCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCG 3715
Db       2124 AGAGCAGTGCTGACTGGCGGGAGGTGTGGGCGCTTGTGCAGGCCGGCATCCGGGATGGGG 2183

Qy       3716 TGTTCACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGA 3775
Db       2184 TGGTACGGCCCCTCAAGTGCACGGTGTTCCATGGGGCCCAGGTGGAGGACGCCTTCCGCT 2243

Qy       3776 GCATGGCGCAGGCGCAG 3792
Db       2244 ACATGGCCCAAGGGAAG 2260
```

```
RESULT 10
BC014631
LOCUS       BC014631               4185 bp    mRNA    linear   HTC 12-SEP-2002
DEFINITION  Homo sapiens, fatty acid synthase, clone IMAGE:3351370, mRNA.
ACCESSION   BC014631
VERSION     BC014631.1  GI:17939526
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4185)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 11 Row: m Column: 1
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1..4185
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3351370"
                     /tissue_type="Eye, retinoblastoma"
                     /clone_lib="NIH_MGC_16"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
BASE COUNT      704 a   1421 c   1324 g    736 t
ORIGIN

  Query Match              2.9%;  Score 123;  DB 11;  Length 4185;
  Best Local Similarity   50.2%;  Pred. No. 7.8e-12;
  Matches  340;  Conservative    0;  Mismatches  325;  Indels   12;  Gaps    1;

Qy       3128 TTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCGCTCTCGG 3187
Db       1593 TGGCCACCTCTGTCCTGCTGTCACCGGACTTCCTCTGGGATGTGCCTTCCAACTGGACGC 1652

Qy       3188 CGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTCGACAGAA 3247
Db       1653 TGGAGGAGGCGGCCTCGGTGCCTGTCGTCTACAGCACGGCCTACTACGCGCTGGTGGTGC 1712

Qy       3248 TAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGGGTCGGTC 3307
Db       1713 GTGGGCGGGTGCGCCCCGGGGAGACGCTGCTCATCCACTCGGGCTCGGGCGGCGTGGGCC 1772

Qy       3308 TCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCCGGCACGC 3367
Db       1773 AGGCCGCCATCGCCATCGCCCTCAGTCTGGGCTGCCGCGTCTTCACCACCGTGGGGTCGG 1832

Qy       3368 CCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAG-----------CG 3415
Db       1833 CTGAGAAGCGGGCGTACCTCCAGGCCAGGTTCCCCCAGCTCGACAGCACCAGCTTCGCCA 1892

Qy       3416 ATTCCCGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAG 3475
Db       1893 ACTCCCGGGACACATCCTTCGAGCAGCATGTGCTGTGGCACACGGGCGGGAAGGGCGTTG 1952

Qy       3476 ACGTCGTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGAT 3535
Db       1953 ACCTGGTCTTGAACTCCTTGGCGGAAGAGAAGCTGCAGGCCAGCGTGAGGTGCTTGGCTA 2012

Qy       3536 CGCACGGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGC 3595
Db       2013 CGCACGGTCGCTTCCTGGAAATTGGCAAATTCGACCTTTCTCAGAACCACCCGCTCGGCA 2072

Qy       3596 TGCGGCCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCG 3655
Db       2073 TGGCTATCTTCCTGAAGAACGTGACATTCCACGGGGTCCTACTGGATGCGTTCTTCAACG 2132

Qy       3656 AGCGGCCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCG 3715
Db       2133 AGAGCAGTGCTGACTGGCGGGAGGTGTGGGCGCTTGTGCAGGCCGGCATCCGGGATGGGG 2192

Qy       3716 TGTTCACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGA 3775
Db       2193 TGGTACGGCCCCTCAAGTGCACGGTGTTCCATGGGGCCCAGGTGGAGGACGCCTTCCGCT 2252

Qy       3776 GCATGGCGCAGGCGCAG 3792
Db       2253 ACATGGCCCAAGGGAAG 2269

RESULT 11
BQ069153
LOCUS       BQ069153               1061 bp    mRNA    linear   EST 02-APR-2002
DEFINITION  AGENCOURT_6738542 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5803443
            5', mRNA sequence.
ACCESSION   BQ069153
VERSION     BQ069153.1  GI:19898199
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1061)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2039  row: h  column: 04
            High quality sequence stop: 563.
FEATURES             Location/Qualifiers
     source          1..1061
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5803443"
                     /tissue_type="neuroblastoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_47"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      219 a    333 c    316 g    193 t
ORIGIN

  Query Match             2.8%;  Score 118.8;  DB 13;  Length 1061;
  Best Local Similarity   51.1%;  Pred. No. 3.8e-11;
  Matches  339;  Conservative    0;  Mismatches  312;  Indels   13;  Gaps    2;

Qy       3193 GAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTCGACAGAATAGCC 3252
              |||||||| |  |||| ||||  |||    ||||| |  ||||||| |          |
Db         40 GAGGCGGCCTCGGTGCCTGTCGTCTACAGCACGGCCTACTACGCGCTGGTGGTGCGTGGG 99

Qy       3253 CGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGGGTCGGTCTCGCC 3312
               || | |  || |||||| | |||| |||||  |||    | ||||| || || |  |||
Db        100 CGGGTGCGCCCCGGGGAGACGCTGCTCATCCACTCGGGCTCGGGCGGCGTGGGCCAGGCC 159

Qy       3313 GCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCCGGCACGCCCGAG 3372
              || |         | |      ||||     |||      | |||  | |   | || |||
Db        160 GCCATCGCCATCGCCCTCAGTCTGGGCTGCCGCGTCTTCACCACCGTGGGGTCGGCTGAG 219

Qy       3373 AAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAG------------CGATTCC 3420
              || || || |||||  || |    |   |   | |   |  ||             | | |||
Db        220 AAGCGGGCGTACCTCCAGGCCAGGTTCCCCCAGCTCGACAGCACCAGCTTCGCCAACTCC 279

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
                ||          ||||     | ||||     |   |||||||  |||| || ||| |
Db        280 CGGGACACATCCTTCGAGCAGCATGTGCTGTGGCACACGGGCGGGAAGGGCGTTGACCTG 339

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              || |  |||||   |  |||  |||   | |   |     ||  |   |   | ||    |||||
Db        340 GTCTTGAACTCCTTGGCGGAAGAGAAGCTGCAGGCCAGCGTGAGGTGCTTGGCTACGCAC 399

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              || || ||  || ||||   ||||      ||||  ||     |||  |||||| ||
Db        400 GGTCGCTTCCTGGAAATTGGCAAATTCGACCTTTCTCAGAACCACCCGCTCGGCATGGCT 459

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
                   ||||||   ||   |  | |||     |||    ||     | | | |     |||| |
Db        460 ATCTTCCTGAAGAACGTGACATTCCACGGGGTCCTACTGGATGCGTTCTTCAACGAGAGC 519

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
                ||       |||  | | | | || |     |  |  |    |||  ||  || ||| |
Db        520 AGTGCTGACTGGCGGGAGGTGTGGGCGCTTGTGCAGGCCGGCATCCGGGATGGGGTGGTA 579

Qy       3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGT-GTCGCCGATGCGTTCCGGAGCAT 3779
              || | |     | ||  | |||         ||||   || |   || || |||||     |||
Db        580 CGGCCCCTCAAGTGCACGGTGTTCCATGGGGCCCCAGGTGGAGGACGCCTTCCGCTACAT 639

Qy       3780 GGCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGAT 3839
              ||  |  | | | ||  |||| ||  |||| ||   | ||  ||    |||   | |
Db        640 GGCCCAGGGAACCACATTGGCAAAGTCGTCCTGCAGGTGCTTGCGGAGGAACCCGAAGG 699

Qy       3840 CCGT 3843
               | ||
Db        700 CAGT 703


RESULT 12
BX332425
LOCUS       BX332425                1201 bp    mRNA    linear   EST 02-MAY-2003
DEFINITION  BX332425 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC012YI01 5-PRIME, mRNA sequence.
ACCESSION   BX332425
VERSION     BX332425.1  GI:30335143
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 8574.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DC012AE01QP1&cluster=8574.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DC012AE01QP1.
FEATURES             Location/Qualifiers
     source          1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC012YI01"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      237 a    340 c    362 g    207 t     55 others
ORIGIN

  Query Match             2.8%;  Score 118.2;  DB 13;  Length 1201;
  Best Local Similarity   52.5%;  Pred. No. 4.9e-11;
  Matches  287;  Conservative    0;  Mismatches  253;  Indels    7;  Gaps    1;

Qy        226 CTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTG 285
              ||||  ||| |   |   | || || |||||||||||||  ||    ||       ||| |
Db        337 CTGAAGGACCTGTCTAGGTTTGATGCCTCCTTCTTCGGAGTCCACCCCAAGCAGGCACAC 396

Qy        286 CGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCC 345
                |||||||||    || ||   ||||||| || |   ||  ||  ||   | | |||
Db        397 ACGATGGACCCTCAGCTGCGGCTGCTGCTGGAAGTCACCTATGAAGCCATCGTGGACGGA 456

Qy        346 GCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCG 405
               |  |||   |||     | |||    || ||  | || || || |||            |  || |
Db        457 GGCATCAACCCAGATTCACTCCGAGGAACACACACTGGCGTCT-------GGGTGGGCGT 509

Qy        406 TCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGC 465
                  || | |||| || ||  | | |           |  |||       |||  |    |  ||
Db        510 GACGGCTCTGAGACCTCGGAGGCCCTGAGCCGAGACCCCGAGACACTCGTGGGCTACAGC 569
```

```
Qy        466 GGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGA 525

Db        570 ATGGTGGGCTGCCAGCGAGCGATGATGGCCAACCGGCTCTCCTTCTTCTTCGACTTCAGA 629

Qy        526 GGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCC 585

Db        630 GGGCCCAGCATCGCACTGGACACAGCCTGCTCCTCCAGCCTGATGGCCCTGCAGAACGCC 689

Qy        586 TGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATG 645

Db        690 TACCAGGCCATCCACAGCGGGCAGTGCCCTGCCGCCATCGTGGGGGGCATCAATGTCCTG 749

Qy        646 TTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGC 705

Db        750 CTGAAGCCCAACACCTCCGTGCAGTTCTTGAGGCTGGGGATGCTCAGCCCCGAGGGCACC 809

Qy        706 TGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTG 765

Db        810 TGCAAGCCCTTCGACACAGCGGGGAATGGGTACTGCCGCTCGGAGGGTGTGGTGGCCGTC 869

Qy        766 GTCCTCA 772

Db        870 CTGCTGA 876
```

```
RESULT 13
AG171124
LOCUS       AG171124                1798 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey
            sequence.
ACCESSION   AG171124
VERSION     AG171124.1  GI:16700802
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1798)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector     : pBACe3.6
                 R.Site 1   : EcoRI
                 R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1..1798
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-040F09.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT        0 a     778 c     655 g       0 t     365 others
ORIGIN
```

```
  Query Match             2.7%;  Score 114.6;  DB 29;  Length 1798;
  Best Local Similarity   40.8%;  Pred. No. 2.2e-10;
  Matches  589;  Conservative    0;  Mismatches  830;  Indels   23;  Gaps    5;

Qy       1782 CAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTT 1841

Db        366 CGGCGCCGGCCGCCCGCCNCCGCNNCGGCNCCCCGCCGGCCGCGCCNCGCCGCGCCGGCC 425

Qy       1842 CGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGG 1901

Db        426 GCCCCCCGCGCNCGCCGCCCCGCCGCGCCCCCCGCCCCCGGNNGGGGNNNGGGGGNGGG 485

Qy       1902 CCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGC 1961

Db        486 GGGNGGCGNNNNNCGGNNNNNNGGGGGCGCGGGGGNGGGCGGGGGNNNNGNGNGGNNNGG 545

Qy       1962 GGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGAT 2021

Db        546 NGGGNCGCGGGGNGGCGCGNNNGGGGGGGNGGGGCGNGCGNGG--GGGGGGGGGGNNNGG 603

Qy       2022 GGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTC 2081

Db        604 GGNGGCGGCGCGCGCGGGCGCGCGCGGCGCGCGGGGNGCGCGGGCGCCGCGCGCGGCGCG 663

Qy       2082 GGTGTCGATCGCCGCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACC 2141

Db        664 GGCGCGGCCGCGGGCGCCCGCGGCGCGCGCGGNCGNGGCGCGCGCGCGCGCGCGCGCG 723

Qy       2142 CGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGT 2201

Db        724 CGCGCGCGCG-----CGGCGCGGCGCGCGCGCGCGCGNCGCGCNCGCGCGCGCCGGNCGC 778

Qy       2202 CTCGCATGCGTTCCACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGC 2261

Db        779 CCGCCGCGGCGCCCGCNCNCCGCGCGCCGCCGCGGCGCNNNGCGCGCCGCGCNCGCGCCG 838

Qy       2262 CGAGTCG-----------GTGAGCTACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCT 2309

Db        839 CGCGNCGCCCCCGCGGCGCGCNCGCCCGCGNCGCCGCGCGCGCGCNCGCCGCGCCGCGCGCC 898

Qy       2310 GAGCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCG 2369

Db        899 NCGCGCCCGCGCGNNCCGCGGCGCCGCGCCCGCCNCGCCCGCNCGGCGCNCNNCGCCGCC 958

Qy       2370 AGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCACCTT 2429

Db        959 CCGCGCCNCNCGCCGCCGCGGCGCNCGCCGCCCGCGCGCCGCCGCCGCCCNCGGCGCGCC 1018

Qy       2430 CGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACGC 2489

Db       1019 NGCNGCNCGCCCGCGCGCCNNNNNNGCGCCGGCGCGCGCGCGCGNGCGCGCGCGCGCNNN 1078

Qy       2490 CCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGA 2549

Db       1079 CGCGCCCGCGNCNCNCCGCGNCGCCCGCCCGCGCGCGCGNCCGCGGGCCGCCNCNCGCCC 1138

Qy       2550 GGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTC 2609

Db       1139 GCCGCGCNNCCGCNCNGCGGCGGCGCGCGCGCGCGCGCGNCGGCGCGCGCCGGCGCGCNC 1198

Qy       2610 AGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGA 2669

Db       1199 GCGGCNNCGCCCGCCGGCGCCGCCGCGCCGCGCGGCGNCGCGCGCGGGCGCGCNCGCGCC 1258

Qy       2670 CACGAAAGCCGACGACGCGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGA 2729

Db       1259 CGGCNGNGCGCGCNCCCGCNCGCGCCGGCCGCGCCCCCGCGCGCNCGCGGCGCGCGCGCG 1318

Qy       2730 GGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCGGCGCAGCGCTCGGCTCGACCATCC 2789

Db       1319 CGCGGGCCGCGCGCGCGCGCGGCGCGGGGCGCCGGCGCGCGCGGCGCGCGGCGCCGCGCGGC 1378

Qy       2790 GCCGCCCGAGAGCGGACGCCGGGAGAAGGTCGAGGCCGCCGGCGACCGTCCGTTCCGGCT 2849
```

```
Db        1379 GCCGCGGGGNCGCCGCNCCGGCGCCGCGCGCGCCNCCGCGCNCGGNCCGCGGCGGNCGNC 1438

Qy        2850 CGAGATCGATGAGCCAGGCGTGCTTGATCACCTCGTGCTTCGGGTCACGGAGCGGCGCGC 2909

Db        1439 CGGCCGCCNCNCGGCNCGCGCGGCGCCGCGGCGCGCGCCGGCGCCNCGCCGCNCCNCN 1498

Qy        2910 CCCTGGTCTGGGCGAGGTCGAGATCGCCGTCGACGCGGCGGGGCTCAGCTTCAATGATGT 2969

Db        1499 CCGCGCCCGCNNCNCNNCCGNNCCGCGCGCCNNCGCGNCCGCGNCNCGCNNCNCNCCGCG 1558

Qy        2970 CCAGCTCGCGCTGGGCATGGTGCCCGACGACCTGCCGGGAAAGCCCAACCCTCCGCTGCT 3029

Db        1559 CGGCGGCGCGCCGGCGCCGCCGGCGCCCGNCCGCNCGG---CGCNCCGCCCCGCGCGNCN 1615

Qy        3030 GCTCGGAGGCGAGTGCGCCGGGCGCATCGTCGC-CGTGGGCGAGGGCGTGAACGGCCTCG 3088

Db        1616 CGCGCGNCGNNCGCGCGCGCGCCGCCGCCGCGCGCGCGGCGNCGGCCGGCCGCNCGCGCG 1675

Qy        3089 TGGTGGGCCAACCGGTCATCGCCCTTTCGGCGGGAGCGTTTGCTACCCACGTCACCACGT 3148

Db        1676 CGCGCCNCCGCGCCGCCGCCGCCCGNCGNCCCGGCNCCGCNNGGCGCCGCCNNCCCCCGG 1735

Qy        3149 CGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCGCTCTCGGCGATCGAGGCGGCCGCCATGC 3208

Db        1736 GCGCCGCGCCGCGCGNCGCGGCGCCCCGCCCGCCCCGGCCGCGCGCGGCCCCGGCCNCNC 1795

Qy        3209 CC 3210

Db        1796 GC 1797
```

RESULT 14
BZ560333

```
LOCUS       BZ560333                 599 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_2364.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_2364, genomic survey sequence.
ACCESSION   BZ560333
VERSION     BZ560333.1  GI:27178168
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..599
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_2364"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      89 a    184 c    199 g    126 t      1 others
ORIGIN

  Query Match             2.7%;  Score 114.4;  DB 29;  Length 599;
  Best Local Similarity   52.2%;  Pred. No. 2.2e-10;
  Matches  253;  Conservative    0;  Mismatches  232;  Indels    0;  Gaps    0;

Qy         34 GATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCGATCTGAGC 93

Db        111 GAACCCGTCGCAATCATCGGCAGCGCTTGCCGTTTTCCCGGCGGAGCGTCGTCGCCGGAG 170

Qy         94 GGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCGCCGAACGC 153

Db        171 GCATTCTGGAACCTCCTTCGTCGTGGGACCGATTTCGTCGGGCCGGTCCCCGCGGGCCGC 230

Qy        154 TGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGACGCCCGTTACG 213

Db        231 TGGGATGGCGAGGCGATTCACGACGCTGGTGCGAGCCGTCCCGGTACCACCTATTGCCGG 290

Qy        214 CGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCATCTCGCCT 273

Db        291 GAGGGGGCCTTCCTTGATGAGGTGGACCGCTTCGACGCGCGTTTCTTCGGGATATCCGCT 350

Qy        274 CGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCTGGGAGGCG 333

Db        351 TCCGAGGCTCGCGAAATGGACCCCCAGCAGCGGCTGCTGCTGGAGGTCGTCTGGGAATGC 410

Qy        334 CTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAGTGTTCATC 393

Db        411 ATGGAACGGGCGGGGCTCTGCTCTGAACGCCTGAAAGGTTCCCGCACTGCGGTCTACGTC 470

Qy        394 GGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCGCAGAGATC 453

Db        471 GGCATGCTGGGCATGGATTACCTTACCCTGCATGCGCGCGAGGCCGGTATCCCGGCGATC 530

Qy        454 GACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCTCGTATGCC 513

Db        531 AATCCATTCTATGCGGCTGGANGGGAGTTCAGCTTCGCCGCCGGGCGCATCGCCTATCAC 590

Qy        514 CTCGG 518

Db        591 CTTGG 595
```

RESULT 15
BU528700

```
LOCUS       BU528700                 899 bp    mRNA    linear   EST 13-SEP-2002
DEFINITION  AGENCOURT_10183164 NIH_MGC_101 Homo sapiens cDNA clone
            IMAGE:6538675 5', mRNA sequence.
ACCESSION   BU528700
VERSION     BU528700.1  GI:22839141
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 899)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2703  row: j  column: 19
            High quality sequence stop: 600.
FEATURES             Location/Qualifiers
     source          1..899
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6538675"
                     /tissue_type="epidermoid carcinoma, cell line"
```

```
                    /lab_host="DH10B (phage-resistant)"
                    /clone_lib="NIH_MGC_101"
                    /note="Organ: lung; Vector: pOTB7; Site_1: EcoRI; Site_2:
                    XhoI; cDNA made by oligo-dT priming. Directionally cloned
                    into EcoRI/XhoI sites using the following 5' adaptor:
                    GGCACGAG(G). Library constructed by Ling Hong in the
                    laboratory of Gerald M. Rubin (University of California,
                    Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                    Superscript II RT (Life Technologies). Note: this is a
                    NIH_MGC Library."
BASE COUNT      145 a    280 c    307 g    166 t      1 others
ORIGIN

  Query Match             2.7%;  Score 113.4;  DB 13;  Length 899;
  Best Local Similarity   49.4%;  Pred. No. 3.3e-10;
  Matches  358;  Conservative    0;  Mismatches  352;  Indels   14;  Gaps    2;

Qy       3128 TTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCGCTCTCGG 3187

Db         46 TGGCCACCTCTGTCCTGCTGTCACCGGACTTCCTCTGGGATGTGCCTTCCAACTGGACGC 105

Qy       3188 CGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTCGACAGAA 3247

Db        106 TGGAGGAGGCGGCCTCGGTGCCTGTCGTCTACAGCACGGCCTACTACGCGCTGGTGGTGC 165

Qy       3248 TAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGGGTCGGTC 3307

Db        166 GTGGGCGGGTGCGCCCCGGGGAGACGCTGCTCATCCACTCGGGCTCGGCGGCGTGGGCC 225

Qy       3308 TCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCCGGCACGC 3367

Db        226 AGGCCGCCATCGCCATCGCCCTCAGTCTGGGCTGCCGCGTCTTCACCACCGTGGGGTCGG 285

Qy       3368 CCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAG------------CG 3415

Db        286 CTGAGAAGCGGGCGTACCTCCAGGCCAGGTTCCCCCAGCTCGACAGCACCAGCTTCGCCA 345

Qy       3416 ATTCCCGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAG 3475

Db        346 ACTCCCGGGACACATCCTTCGAGCAGCATGTGCTGTGGCACACGGGCGGGAAGGGCGTTG 405

Qy       3476 ACGTCGTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGAT 3535

Db        406 ACCTGGTCTTGAACTCCTTGGCGGAAGAGAAGCTGCAGGCCAGCGTGAGGTGCTTGGCTA 465

Qy       3536 CGCACGGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGC 3595

Db        466 CGCACGGTCGCTTCCTGGAAATTGGCAAATTCGACCTTTCTCAGAACCACCCGCTCGGCA 525

Qy       3596 TGCGGCCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCG 3655

Db        526 TGGCTATCTTCCTGAAGAACGTGACATTCCACGGGGTCCTACTGGATGCGTTCTTCAACG 585

Qy       3656 AGCGGCCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCG 3715

Db        586 AGAGCAGTGCTGACTGGCGGGAGGTGTGGGCGCTTGTGCAGGCCGGCATCCGGGATGGGG 645

Qy       3716 TGTTCACCCCTCCCCCCATCGCGACGCTCCC--GATCGCCCGTGTCGCCGATGCGTTCCG 3773

Db        646 TGGTACGGGCCCTCAAGTGCACGGTGTTCCCATGGGGCCCAGGTGGGAGGACGCCTTCCG 705

Qy       3774 GAGCATGGCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGT 3833

Db        706 CTACCTGNCCCAGGGGAAACACATTGGCAAAGTCGCCGTGCAGGGGCCTTGCCGGAGGAA 765

Qy       3834 CCAG 3837

Db        766 CCCG 769

Search completed: October  3, 2003, 17:47:55
Job time : 8435.03 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:         October  4, 2003, 03:31:13 ; Search time 981.368 Seconds
                                             (without alignments)
                                             11088.586 Million cell updates/sec

Title:          US-10-014-717-1_COPY_7610_11875
Perfect score:  4266
Sequence:       1 gtggcggatcgtcccatcga..........gggaaatcattgccctatga 4266

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters:      3416838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      Published_Applications_NA:*
                1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
                12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                13:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                14:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                15:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                16:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                17:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4266	100.0	68750	13	US-10-014-717-1	Sequence 1, Appli
	2	814.2	19.1	3978	12	US-09-953-348-19	Sequence 19, Appl
	3	814.2	19.1	3978	14	US-10-267-255-19	Sequence 19, Appl
	4	814.2	19.1	12249	12	US-09-953-348-74	Sequence 74, Appl
	5	814.2	19.1	12249	14	US-10-267-255-74	Sequence 74, Appl
	6	814.2	19.1	18331	12	US-09-953-348-96	Sequence 96, Appl
	7	814.2	19.1	18331	14	US-10-267-255-96	Sequence 96, Appl
	8	770.8	18.1	5355	14	US-10-205-032-19	Sequence 19, Appl
	9	770.8	18.1	60196	14	US-10-205-032-1	Sequence 1, Appli
	10	765.8	18.0	6360	14	US-10-212-962-2	Sequence 2, Appli
	11	761.2	17.8	10839	14	US-10-156-761-2882	Sequence 2882, Ap
	12	761.2	17.8	125746	14	US-10-156-761-15102	Sequence 15102, A
c	13	761.2	17.8	9025608	14	US-10-156-761-1	Sequence 1, Appli
	14	742	17.4	15872	10	US-09-861-289-1	Sequence 1, Appli
	15	742	17.4	15872	10	US-09-860-846-1	Sequence 1, Appli
	16	742	17.4	15872	11	US-09-988-384B-1	Sequence 1, Appli
	17	742	17.4	15872	11	US-09-836-821-1	Sequence 1, Appli
	18	732.8	17.2	18435	14	US-10-156-761-412	Sequence 412, App
	19	732.8	17.2	100000	14	US-10-156-761-15103	Sequence 15103, A
	20	715.8	16.8	18438	14	US-10-156-761-2886	Sequence 2886, Ap
	21	707	16.6	23238	14	US-10-156-761-415	Sequence 415, App
	22	695.4	16.3	11817	14	US-10-156-761-2884	Sequence 2884, Ap
	23	679.6	15.9	10692	14	US-10-156-761-414	Sequence 414, App
	24	679.4	15.9	11238	14	US-10-205-032-15	Sequence 15, Appl
	25	669.2	15.7	14427	14	US-10-156-761-1540	Sequence 1540, Ap
	26	667.6	15.6	5505	14	US-10-156-761-413	Sequence 413, App
	27	666	15.6	14520	14	US-10-156-761-2885	Sequence 2885, Ap
	28	655.6	15.4	10056	14	US-10-156-761-411	Sequence 411, App
	29	652.6	15.3	11910	14	US-10-156-761-2879	Sequence 2879, Ap
c	30	652.6	15.3	125746	14	US-10-156-761-15102	Sequence 15102, A
	31	652.6	15.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
	32	648.6	15.2	4725	14	US-10-205-032-17	Sequence 17, Appl
	33	633	14.8	77536	12	US-09-940-316B-1	Sequence 1, Appli
	34	627.8	14.7	13494	14	US-10-156-761-2355	Sequence 2355, Ap
	35	627.6	14.7	13842	10	US-09-861-289-30	Sequence 30, Appl
	36	627.6	14.7	13842	10	US-09-860-846-30	Sequence 30, Appl
	37	627.6	14.7	13842	11	US-09-988-384B-30	Sequence 30, Appl
	38	627.6	14.7	13842	11	US-09-836-821-30	Sequence 30, Appl
	39	627.6	14.7	36778	10	US-09-861-289-5	Sequence 5, Appli
	40	627.6	14.7	36778	10	US-09-860-846-5	Sequence 5, Appli
	41	627.6	14.7	36778	11	US-09-836-821-5	Sequence 5, Appli
	42	627.6	14.7	37948	11	US-09-988-384B-5	Sequence 5, Appli
	43	627.6	14.7	38506	11	US-09-793-708-19	Sequence 19, Appl
	44	627.6	14.7	38506	12	US-10-201-365-1	Sequence 1, Appli
	45	627.6	14.7	38506	12	US-10-160-539-19	Sequence 19, Appl

ALIGNMENTS

```
RESULT 1
US-10-014-717-1
; Sequence 1, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/10/014,717
;  CURRENT FILING DATE:  2001-11-13
;  PRIOR APPLICATION NUMBER: US/09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-10-014-717-1

  Query Match            100.0%;  Score 4266;  DB 13;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGAGCTCCCAAGAAATCGTGCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGAGCTCCCAAGAAATCGTGCTG 8509

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy       1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy       1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy       1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy       2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy       2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy       2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy       2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy       2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949
```

```
Qy       2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875
```

RESULT 2
US-09-953-348-19
; Sequence 19, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingqing
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953,348

```
;  CURRENT FILING DATE:  2001-09-12
;  PRIOR APPLICATION NUMBER: PCT/US00/06394
;  PRIOR FILING DATE: 2000-03-10
;  PRIOR APPLICATION NUMBER: 09/266965
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 153
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19
;   LENGTH: 3978
;   TYPE: DNA
;   ORGANISM: Streptomyces lavendulae
US-09-953-348-19

  Query Match             19.1%;  Score 814.2;  DB 12;  Length 3978;
  Best Local Similarity   57.6%;  Pred. No. 7.7e-170;
  Matches 1549;  Conservative    0;  Mismatches 1108;  Indels   30;  Gaps    4;

Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77
Db         90 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 149

Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137
Db        150 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 209

Qy        138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194
Db        210 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 269

Qy        195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254
Db        270 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 329

Qy        255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314
Db        330 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 389

Qy        315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374
Db        390 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 449

Qy        375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434
Db        450 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 509

Qy        435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGC 494
Db        510 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 563

Qy        495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554
Db        564 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 623

Qy        555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614
Db        624 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 683

Qy        615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674
Db        684 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 743

Qy        675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734
Db        744 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 803

Qy        735 GTTCGGACGAGGCGAAGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794
Db        804 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 863

Qy        795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854
Db        864 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 923

Qy        855 CAGCGGTCTGACCGTGCCGAACGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914
Db        924 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 983

Qy        915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974
Db        984 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 1043

Qy        975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034
Db       1044 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGCCCGCGGCGA 1103

Qy       1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094
Db       1104 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 1163

Qy       1095 GGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154
Db       1164 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 1223

Qy       1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214
Db       1224 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 1283

Qy       1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274
Db       1284 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 1343

Qy       1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334
Db       1344 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCCGCCCGCCGAGGAGGCACC 1403

Qy       1335 ACCGCCG------------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376
Db       1404 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 1463

Qy       1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436
Db       1464 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 1523

Qy       1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496
Db       1524 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 1583

Qy       1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556
Db       1584 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 1643

Qy       1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616
Db       1644 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 1703

Qy       1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCT 1676
Db       1704 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCAGCAGCCCGGCATGGGCCGCGAACT 1763

Qy       1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db       1764 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 1823

Qy       1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db       1824 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 1883

Qy       1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db       1884 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 1943

Qy       1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db       1944 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2003

Qy       1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
```

```
Db       2004 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 2063

Qy       1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db       2064 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 2123

Qy       2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db       2124 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 2183

Qy       2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
Db       2184 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 2243

Qy       2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216
Db       2244 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 2303

Qy       2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276
Db       2304 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 2363

Qy       2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336
Db       2364 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 2423

Qy       2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396
Db       2424 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 2483

Qy       2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456
Db       2484 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 2543

Qy       2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513
Db       2544 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 2603

Qy       2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGT 2573
Db       2604 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 2663

Qy       2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCC 2633
Db       2664 CGGCCTCTCGCCGGCGTGGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 2723

Qy       2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680
Db       2724 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 2770
```

RESULT 3
US-10-267-255-19
; Sequence 19, Application US/10267255
; Publication No. US20030124689A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19
;   LENGTH: 3978
;   TYPE: DNA
;   ORGANISM: Streptomyces lavendulae
US-10-267-255-19

```
  Query Match             19.1%;  Score 814.2;  DB 14;  Length 3978;
  Best Local Similarity   57.6%;  Pred. No. 7.7e-170;
  Matches 1549;  Conservative    0;  Mismatches 1108;  Indels   30;  Gaps    4;

Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77
Db         90 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 149

Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137
Db        150 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 209

Qy        138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194
Db        210 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 269

Qy        195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254
Db        270 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 329

Qy        255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314
Db        330 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 389

Qy        315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374
Db        390 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 449

Qy        375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434
Db        450 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 509

Qy        435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGC 494
Db        510 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 563

Qy        495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554
Db        564 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 623

Qy        555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614
Db        624 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 683

Qy        615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674
Db        684 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 743

Qy        675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734
Db        744 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 803

Qy        735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794
Db        804 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 863

Qy        795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854
Db        864 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 923

Qy        855 CAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914
Db        924 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 983
```

```
Qy        915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974
Db        984 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 1043
Qy        975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034
Db       1044 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGGCCCGCGGCGA 1103
Qy       1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094
Db       1104 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 1163
Qy       1095 GGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154
Db       1164 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 1223
Qy       1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214
Db       1224 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 1283
Qy       1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274
Db       1284 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 1343
Qy       1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334
Db       1344 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 1403
Qy       1335 ACCGCCG-----------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376
Db       1404 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 1463
Qy       1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436
Db       1464 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 1523
Qy       1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496
Db       1524 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 1583
Qy       1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556
Db       1584 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 1643
Qy       1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616
Db       1644 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 1703
Qy       1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGCCGTGGGCT 1676
Db       1704 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCCGGCATGGGCCGCGAACT 1763
Qy       1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db       1764 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 1823
Qy       1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db       1824 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 1883
Qy       1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db       1884 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 1943
Qy       1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db       1944 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2003
Qy       1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
Db       2004 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 2063
Qy       1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db       2064 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 2123
Qy       2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db       2124 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 2183
Qy       2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
Db       2184 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 2243
Qy       2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216
Db       2244 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 2303
Qy       2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276
Db       2304 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 2363
Qy       2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336
Db       2364 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 2423
Qy       2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396
Db       2424 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 2483
Qy       2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456
Db       2484 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 2543
Qy       2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513
Db       2544 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 2603
Qy       2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGT 2573
Db       2604 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 2663
Qy       2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCC 2633
Db       2664 CGGCCTCTCGCCGGCGTGGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACTGCC 2723
Qy       2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680
Db       2724 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 2770
```

RESULT 4
US-09-953-348-74
; Sequence 74, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingqing
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953,348
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: PCT/US00/06394
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/266965
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 12249
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-953-348-74

```
Query Match             19.1%;  Score 814.2;  DB 12;  Length 12249;
Best Local Similarity   57.6%;  Pred. No. 7e-170;
Matches 1549;  Conservative    0;  Mismatches 1108;  Indels   30;  Gaps    4;

Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77
Db       1077 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 1136

Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137
Db       1137 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 1196

Qy        138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194
Db       1197 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 1256

Qy        195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254
Db       1257 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 1316

Qy        255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314
Db       1317 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 1376

Qy        315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374
Db       1377 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 1436

Qy        375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434
Db       1437 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 1496

Qy        435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGACGATGCCCAGCGTCGGAGC 494
Db       1497 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 1550

Qy        495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554
Db       1551 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 1610

Qy        555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614
Db       1611 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 1670

Qy        615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674
Db       1671 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 1730

Qy        675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734
Db       1731 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 1790

Qy        735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794
Db       1791 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 1850

Qy        795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854
Db       1851 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 1910

Qy        855 CAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914
Db       1911 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 1970

Qy        915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974
Db       1971 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 2030

Qy        975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034
Db       2031 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGCCCGCGGCGA 2090

Qy       1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094
Db       2091 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 2150

Qy       1095 GGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154
Db       2151 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 2210

Qy       1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214
Db       2211 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 2270

Qy       1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274
Db       2271 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 2330

Qy       1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334
Db       2331 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 2390

Qy       1335 ACCGCCG-----------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376
Db       2391 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 2450

Qy       1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436
Db       2451 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 2510

Qy       1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496
Db       2511 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 2570

Qy       1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556
Db       2571 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 2630

Qy       1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616
Db       2631 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 2690

Qy       1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCT 1676
Db       2691 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCCGGCATGGGCCGCGAACT 2750

Qy       1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db       2751 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 2810

Qy       1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db       2811 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 2870

Qy       1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db       2871 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 2930

Qy       1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db       2931 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2990

Qy       1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
Db       2991 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 3050

Qy       1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db       3051 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 3110

Qy       2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db       3111 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 3170

Qy       2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
```

```
Db       3171 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 3230

Qy       2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216

Db       3231 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 3290

Qy       2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276

Db       3291 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 3350

Qy       2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336

Db       3351 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 3410

Qy       2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396

Db       3411 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 3470

Qy       2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456

Db       3471 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 3530

Qy       2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513

Db       3531 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 3590

Qy       2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGCTCTGGGCCGT 2573

Db       3591 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 3650

Qy       2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCC 2633

Db       3651 CGGCCTCTCGCCGGCGTGGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 3710

Qy       2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680

Db       3711 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 3757
```

RESULT 5
US-10-267-255-74
; Sequence 74, Application US/10267255
; Publication No. US20030124689A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 74
; LENGTH: 12249
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-10-267-255-74

```
  Query Match             19.1%;  Score 814.2;  DB 14;  Length 12249;
  Best Local Similarity   57.6%;  Pred. No. 7e-170;
  Matches 1549;  Conservative    0;  Mismatches 1108;  Indels   30;  Gaps    4;

Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77

Db       1077 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 1136

Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137

Db       1137 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 1196

Qy        138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194

Db       1197 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 1256

Qy        195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254

Db       1257 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 1316

Qy        255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314

Db       1317 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 1376

Qy        315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374

Db       1377 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 1436

Qy        375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434

Db       1437 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 1496

Qy        435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGACGATGCCCAGCGTCGGAGC 494

Db       1497 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 1550

Qy        495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554

Db       1551 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 1610

Qy        555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614

Db       1611 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 1670

Qy        615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674

Db       1671 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 1730

Qy        675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734

Db       1731 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 1790

Qy        735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794

Db       1791 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 1850

Qy        795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854

Db       1851 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 1910

Qy        855 CAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914

Db       1911 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 1970

Qy        915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974

Db       1971 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 2030

Qy        975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034

Db       2031 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGCCCGCGGCGA 2090

Qy       1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094

Db       2091 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 2150
```

```
Qy       1095 GGGGATCACTGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGCAGATTCCTGCGCA 1154
Db       2151 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 2210
Qy       1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214
Db       2211 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 2270
Qy       1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGTGAGCTCGTT 1274
Db       2271 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 2330
Qy       1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334
Db       2331 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 2390
Qy       1335 ACCGCCG------------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376
Db       2391 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 2450
Qy       1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436
Db       2451 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 2510
Qy       1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496
Db       2511 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 2570
Qy       1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556
Db       2571 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 2630
Qy       1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616
Db       2631 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 2690
Qy       1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCT 1676
Db       2691 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCCGGCATGGGCCGCGAACT 2750
Qy       1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db       2751 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 2810
Qy       1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db       2811 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 2870
Qy       1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db       2871 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 2930
Qy       1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db       2931 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2990
Qy       1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
Db       2991 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 3050
Qy       1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db       3051 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 3110
Qy       2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db       3111 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 3170
Qy       2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
Db       3171 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 3230
Qy       2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216
Db       3231 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 3290
Qy       2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276
Db       3291 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 3350
Qy       2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336
Db       3351 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 3410
Qy       2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396
Db       3411 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 3470
Qy       2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456
Db       3471 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 3530
Qy       2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513
Db       3531 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 3590
Qy       2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGT 2573
Db       3591 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 3650
Qy       2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCC 2633
Db       3651 CGGCCTCTCGCCGGCGTGGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 3710
Qy       2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680
Db       3711 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 3757
```

RESULT 6
US-09-953-348-96
; Sequence 96, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingqing
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953,348
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: PCT/US00/06394
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/266965
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 18331
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-953-348-96

```
  Query Match             19.1%;  Score 814.2;  DB 12;  Length 18331;
  Best Local Similarity   57.6%;  Pred. No. 6.8e-170;
  Matches 1549;  Conservative    0;  Mismatches 1108;  Indels   30;  Gaps    4;

Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77
Db       1077 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 1136
Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137
Db       1137 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 1196
```

```
Qy     138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194
Db    1197 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 1256

Qy     195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254
Db    1257 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 1316

Qy     255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314
Db    1317 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 1376

Qy     315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374
Db    1377 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 1436

Qy     375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434
Db    1437 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 1496

Qy     435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGC 494
Db    1497 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 1550

Qy     495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554
Db    1551 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 1610

Qy     555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614
Db    1611 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 1670

Qy     615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674
Db    1671 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 1730

Qy     675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTCGGCGGAGGCCGATGG 734
Db    1731 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 1790

Qy     735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794
Db    1791 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 1850

Qy     795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854
Db    1851 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 1910

Qy     855 CAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914
Db    1911 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 1970

Qy     915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974
Db    1971 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 2030

Qy     975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034
Db    2031 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGGCCCGCGGCGA 2090

Qy    1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094
Db    2091 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 2150

Qy    1095 GGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154
Db    2151 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 2210

Qy    1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214
Db    2211 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 2270

Qy    1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274
Db    2271 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 2330

Qy    1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334
Db    2331 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 2390

Qy    1335 ACCGCCG-----------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376
Db    2391 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 2450

Qy    1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436
Db    2451 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 2510

Qy    1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496
Db    2511 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 2570

Qy    1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556
Db    2571 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 2630

Qy    1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616
Db    2631 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 2690

Qy    1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGCCGTGGGCT 1676
Db    2691 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCCGGCATGGGCCGCGAACT 2750

Qy    1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db    2751 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 2810

Qy    1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db    2811 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 2870

Qy    1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db    2871 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 2930

Qy    1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db    2931 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2990

Qy    1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
Db    2991 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 3050

Qy    1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db    3051 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 3110

Qy    2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db    3111 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 3170

Qy    2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
Db    3171 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 3230

Qy    2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216
Db    3231 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 3290

Qy    2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276
Db    3291 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 3350

Qy    2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336
```

```
Db       3351 CAGCGCGCCGTCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 3410

Qy       2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396

Db       3411 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 3470

Qy       2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456

Db       3471 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 3530

Qy       2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513

Db       3531 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 3590

Qy       2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGT 2573

Db       3591 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 3650

Qy       2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCC 2633

Db       3651 CGGCCTCTCGCCGGCGTGGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 3710

Qy       2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680

Db       3711 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 3757
```

RESULT 7
US-10-267-255-96
; Sequence 96, Application US/10267255
; Publication No. US20030124689A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 96
; LENGTH: 18331
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-10-267-255-96

```
  Query Match             19.1%;  Score 814.2;  DB 14;  Length 18331;
  Best Local Similarity   57.6%;  Pred. No. 6.8e-170;
  Matches 1549;  Conservative    0;  Mismatches 1108;  Indels   30;  Gaps    4;

Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77

Db       1077 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 1136

Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137

Db       1137 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 1196

Qy        138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194

Db       1197 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 1256

Qy        195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254

Db       1257 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 1316

Qy        255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314

Db       1317 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 1376

Qy        315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374

Db       1377 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 1436

Qy        375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434

Db       1437 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 1496

Qy        435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGC 494

Db       1497 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 1550

Qy        495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554

Db       1551 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 1610

Qy        555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614

Db       1611 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 1670

Qy        615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674

Db       1671 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 1730

Qy        675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734

Db       1731 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 1790

Qy        735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794

Db       1791 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 1850

Qy        795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854

Db       1851 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 1910

Qy        855 CAGCGGTCTGACCGTGCCGAACGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914

Db       1911 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 1970

Qy        915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974

Db       1971 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 2030

Qy        975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034

Db       2031 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGGCCCGCGGCGA 2090

Qy       1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094

Db       2091 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 2150

Qy       1095 GGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154

Db       2151 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 2210

Qy       1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214

Db       2211 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 2270

Qy       1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274

Db       2271 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 2330
```

```
Qy       1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334
Db       2331 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 2390

Qy       1335 ACCGCCG------------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376
Db       2391 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 2450

Qy       1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436
Db       2451 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 2510

Qy       1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496
Db       2511 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 2570

Qy       1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556
Db       2571 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 2630

Qy       1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616
Db       2631 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 2690

Qy       1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCT 1676
Db       2691 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCGGCATGGGCCGCGAACT 2750

Qy       1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db       2751 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 2810

Qy       1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db       2811 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 2870

Qy       1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db       2871 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 2930

Qy       1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db       2931 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2990

Qy       1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
Db       2991 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 3050

Qy       1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db       3051 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 3110

Qy       2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db       3111 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 3170

Qy       2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
Db       3171 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 3230

Qy       2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216
Db       3231 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGTCAGCCACGCCTTCCA 3290

Qy       2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276
Db       3291 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 3350

Qy       2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336
Db       3351 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 3410

Db       3411 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 3470

Qy       2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456
Db       3471 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 3530

Qy       2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513
Db       3531 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 3590

Qy       2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGCTCTGGGCCGT 2573
Db       3591 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 3650

Qy       2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCC 2633
Db       3651 CGGCCTCTCGCCGGCGTGGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 3710

Qy       2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680
Db       3711 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 3757
```

RESULT 8
US-10-205-032-19
; Sequence 19, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-2US
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
;   LENGTH: 5355
;   TYPE: DNA
;   ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-19

```
  Query Match             18.1%;  Score 770.8;  DB 14;  Length 5355;
  Best Local Similarity   54.3%;  Pred. No. 2.6e-160;
  Matches 1868;  Conservative    0;  Mismatches 1517;  Indels   57;  Gaps   13;

Qy         32 AAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCGATCTGA 91
Db        101 ACGGGCCGATCGCCGTCGTGGGCATGGCCTGCCGCTACCCGGGCGGGGTGTCCTCTCCGG 160

Qy         92 GCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCGCCGAAC 151
Db        161 AGGACCTGTGGCGGCTGGTCGAGACGGGCACGGACGCGATCGGCGGCTTCCCCACCGACC 220

Qy        152 ---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGACGCCCG 208
Db        221 GTGGCTGGGACGTCGACGCCGTGTACGACCCGGATCCTGAGTCGCGGAACACCACCTACT 280

Qy        209 TTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCATCT 268
Db        281 GCCGGGAGGGCGGGTTCCTGGCCGGGGCAGGAGACTTCGACGCCGCGTTCTTCGGGGTGT 340

Qy        269 CGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCTGGG 328
Db        341 CGCCGCACGAGGCCGTGGTCATGGACCCCCAGCAGCGGCTGCTTCTGGAGGTGTCCTGGG 400

Qy        329 AGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAGTGT 388
Db        401 AGGCGCTGGAGCGGTCCGGGACCGACCCGCACAGCCTGCGCGGCTCGCGCACCGGGGTCT 460

Qy        389 TCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCGCAG 448
```

```
Db   461 ACGTCGGTGCGGCCCACCAGGGGTACGCGGTCGACGCCGGTCAGGTGCCGGAGGGCGCGG 520

Qy   449 AGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCTCGT 508
Db   521 AG------GGGTTCCGGCTGACCGGCAGCGCCGACGCCGTCCTGTCCGGACGGATCTCGT 574

Qy   509 ATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGCTGG 568
Db   575 ACCTGCTCGGGCTGGAGGGTCCGGCCCTGACCGTCGAGACGGCCTGCTCGTCCTCGCTGG 634

Qy   569 TGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGGCTG 628
Db   635 TGGCGGTGCACCTCGCGGTGCAGGCGCTGCGCCGGGGCGAGTGCGGGCTGGCACTGGCCG 694

Qy   629 GTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGGCGC 688
Db   695 GCGGGGTCGCCGTGATGCCCGACCCGGCGGCATTCGTGGAGTTCTCCCGGCAGCGGGGCC 754

Qy   689 TGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAGGCG 748
Db   755 TCGCGGCGGACGGGCGCTGCCGGGCGTTCGGGGCGGCGCGGACGGCACCGGCTGGGCGG 814

Qy   749 AAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATCGGA 808
Db   815 AGGGCGTCGGTGTGCTGGTCCTGCAACGGCTCTCCGACGCGGTGCGCGACGGCCGCTGGG 874

Qy   809 TATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTGACCG 868
Db   875 TGCTGGGCGTGATCCGGGGTTCGGCCGTCAACCAGGACGGGGCCAGCAACGGGCTGACCG 934

Qy   869 TGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGCTGCG 928
Db   935 CCCCGAGCGGCCCCGCCCAGCAGCGGGTCATCCGGCAGGCGCTGACCGACGCCCGGCTCG 994

Qy   929 CCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGACCCCA 988
Db   995 GCGCCGACCAGATCGACGCGGTCGAGGCGCACGGCACGGGCACCCGGCTCGGCGACCCGA 1054

Qy   989 TCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCGCTGC 1048
Db  1055 TCGAGGCGCAGGCGCTGATCGCCGCCTACGGCGCCGACCGGACCCCGGACCGGCCGCTCT 1114

Qy  1049 TGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACTGGGC 1108
Db  1115 GGCTCGGCTCGTTGAAGTCGAACATCGGGCACGCCCAGGCGGCGGCCGGCGTCGGCGGCC 1174

Qy  1109 TGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGCGCAGG 1168
Db  1175 TGATCAAGATGCTCCTGGCGATGCGGGCCGGGACGCTCCCACCCACCCTGCACGCCGACG 1234

Qy  1169 CGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCGGACAC 1228
Db  1235 TCCCGACCCCGCTGGTCGACTGGTCCGCCGGTGTCGTCCGGCTGTCGACCGGGGTGGTGC 1294

Qy  1229 CGTGGCCGGACTGG---AATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGAGCG 1285
Db  1295 CCTGGCCCGCGTTGCCCGGGGCGCCCCGCAGGGCCGGGATCTCCGCGTTCGGGGTGAGCG 1354

Qy  1286 GGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCGGCGC 1345
Db  1355 GCACCAACGCGCACGTGATCGTCGAGCAGCCGCCGCCGGTCCCGGTCGACGACCCGGCGC 1414

Qy  1346 C---------GGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCC 1396
Db  1415 CACCCACGAGGACCCTGCCGCTGGTGCCGTGGGTGCTCTCCGGCCGGACGGAGGCGGCGC 1474

Qy  1397 TGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGG 1456
Db  1475 TGCGCGCCCAGGCGGACCGGTTGCGTACGCACCTGGCGGCGCACCCCGACGCGGACCCGC 1534

Qy  1457 GCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGG 1516
Db  1535 TGGACGTGGGATTCTCCCTGGCCACCAGCCGGGCCGCGCTGGAGCACCGGGCCGTGCTGG 1594

Qy  1517 CGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGT 1576
Db  1595 TGGCCGCCGACCGCGACGGCCTGCTCCGCCTCGTCGACGCGCTGGCCGCCGGCGAGCCGG 1654

Qy  1577 CGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAGCTCGCCTTTCTCTTCA 1636
Db  1655 CGGCGGGCCTGATCCGGGGCACGGTACGTCACGATCGCCG---GACCGGGTTCCTCTTCG 1711

Qy  1637 CCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGT 1696
Db  1712 CCGGGCAGGGCGGCCAGCGCGTCGGGATGGCGCGCGAACTGTACGAGGCGTTCCCCGCCT 1771

Qy  1697 TCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCC 1756
Db  1772 TCGCCGACGCCCTGGACCAGCTCGCCGCCCGGCTGGACCGGCACCTCGATCGTCCGCTGC 1831

Qy  1757 GCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCT 1816
Db  1832 TGCGGGTGCTGTTCGCCGAGCCGGGGTCGGACGACGCCCGGCTGCTCGACGGCACCCGGT 1891

Qy  1817 TCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGG 1876
Db  1892 ACGCGCAGGCCGCCCTCTTCGCCGTCGAGGTGGCGTTGTTCCGACTGGTCCACGGCTGGG 1951

Qy  1877 GTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGG 1936
Db  1952 GGGTCCGGGCCGACGTGCTGCTCGGCCACTCGGTGGGCGAGCTGGCGGCCGCGCACGTGG 2011

Qy  1937 CGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGC 1996
Db  2012 CCGGCGTACTCGACGTGGACGACGCGTGCGAGCTGGTCGCGGCGCGGGGCCGGCTGATGG 2071

Qy  1997 AGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTG 2056
Db  2072 GGGAGCTGCCGTCGGGCGGCGCGATGGTGGCGGTCCGGGCCACCGAGGAGGAGGTCGGGC 2131

Qy  2057 CTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCCGGACCAGG 2116
Db  2132 CCCTGCTCG------ACGGGCAGCGGGTCGCGGTGGCGGCGGTCAACGGCCCGCGCTCGG 2185

Qy  2117 TGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCG 2176
Db  2186 TCGTGGTCTCCGGCGACGAGGAGGCGGTGCTGGCCGTGGCCGCCCGGTGCGCCGCCCTCG 2245

Qy  2177 GGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGA 2236
Db  2246 GCCACCGGACGCGACGCCTCAACGTCAGCCACGCGTTCCACTCCCCGCACGTGGAGGCGA 2305

Qy  2237 TGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCC 2296
Db  2306 TGCTGGAGCCGTTCCGGCGGGTGGCGCGGGGCCTGACGTACCATGCCCCGACGATCCCGG 2365

Qy  2297 TGGTCAGCAATCTGA---GCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATT 2353
Db  2366 TGGTGTCGAACGCGACGGGCCGGCTCGCCACCGCCGACGCGCTGCGCGACCCCGGTTACT 2425

Qy  2354 GGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGG 2413
Db  2426 GGGTCCGGCACGTCCGCCAGCCGTCCGGTTCCGGGACGGGGTGCGGGCCGCCCGCGACC 2485

Qy  2414 CCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCTGGTGCCTG 2473
Db  2486 AGGGGGCCACCGCCTTCGTCGGGCTCGGCCCGGACGGGGTGCTGTGCGCGTTGGCCGAGG 2545

Qy  2474 CCTGCAT---GCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGACG 2530
Db  2546 AGTGCCTCGGGCCCACCGGCGACGTGCTGCTGCTGCCGGTGCTGCGCCCCGGTCGGCCGG 2605

Qy  2531 AGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCCT 2590
Db  2606 AGCCCGCCACCCTGCTGGCCGCCCTGGCCGGGGCGTACGCCGGCGGCGCGGAAATGGACT 2665
```

```
Qy       2591 GGCCGGCCTCTTCCC---CTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGC 2647
Db       2666 GGTCCCGGGTGTTCGCGGGCACCGGCGCGCGCAGGGTCGAGCTGCCCACGTACGCCTTCC 2725

Qy       2648 AGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGACCGCCGTG 2707
Db       2726 AGCACCGGCGCTACTGGCTGGCGCCGGGCCCGCCGTCGGCCCGCCGCGACGACGCCTG-- 2783

Qy       2708 CTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCGGC 2767
Db       2784 --GCGGTACCGGATCGCCTGGCGGCCCCTGCCGACCGTGCCCGCCGCCGCCGGGACCGAG 2841

Qy       2768 GCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTCG-AGGCC 2826
Db       2842 ACGGTGGCCGGGGCGTGGTTGCTGGTGGTCCCCGCCCACGACGGCGTCGCGTCGCTCGCC 2901

Qy       2827 GCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCACCTCGTG 2886
Db       2902 GACGCCGCCGAGCGGGCCGTGCACCGGGGCGGGGCCACGGTCACCCGGCTGACGGTGGAC 2961

Qy       2887 CTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTCGACGCG 2946
Db       2962 GCCGCCGACGTGGACCGGGACACCCTCGCCGCCGTGCTGACCGAGGCCGCCGCCGACGCG 3021

Qy.      2947 GCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGACCTGCCG 3006
Db       3022 GACGGC------------GGGCCGGACGGGGTGCTCTGCCTGCTGGGCCTCGACGACCGG 3069

Qy       3007 GGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTCGCCGTG 3066
Db       3070 GCACATCCCCGGTCCGCCTCGGTGCCCCGCGGGGTGCTGGCGACCCTGTCCCTCGCCCAG 3129

Qy       3067 GGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCGGGAGCG 3126
Db       3130 GCCCTGACCGACCTGGGGGCCTCCGCGCGGCTGTGGTGCGTGACCCGGGGGGCGGTCGCC 3189

Qy       3127 TTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCGCTCTCG 3186
Db       3190 GTGACGCCCGGCGAGTCCCCGTCGGTCGCCGGAGCCCAGTTGTGGGGCTTCGGCCGCGTG 3249

Qy       3187 GCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTCGACAGA 3246
Db       3250 GCCGCGCTCGAACTCCCCCGGTCCTGGGGCGGCCTGGTGGACCTGCCGGTCGACCCGGAC 3309

Qy       3247 ATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGGGTCGGT 3306
Db       3310 GACCGGGACTGGGACCTGCTGCGGCGCGCGCTGCGCGGCCCGGAGGACCAGGTCGCGGTC 3369

Qy       3307 CTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCCGGCACG 3366
Db       3370 CGGGGGGCGGTCGGGTACGCCCGGCGGCTGGTCCCCG-CGCCCGCGCCCCGGGCCGAGCG 3428

Qy       3367 CCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCCCGCTCG 3426
Db       3429 GGCCTGGCGTCCGCGCGGCACGGTCCTGGTGACCGGCGGTACGGGCGCGCTCGGCGCGCA 3488

Qy       3427 GACCGGTTCGTCGCCGACGTGC 3448
Db       3489 CACGGCCCGCTGGCTGGCGCGC 3510
```

RESULT 9
US-10-205-032-1
; Sequence 1, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-2US
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 60196
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-1

```
  Query Match             18.1%;  Score 770.8;  DB 14;  Length 60196;
  Best Local Similarity   54.3%;  Pred. No. 2.1e-160;
  Matches 1868;  Conservative    0;  Mismatches 1517;  Indels   57;  Gaps   13;

Qy         32 AAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCGATCTGA 91
Db      41149 ACGGGCCGATCGCCGTCGTGGGCATGGCCTGCCGCTACCCGGGCGGGGTGTCCTCTCCGG 41208

Qy         92 GCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCGCCGAAC 151
Db      41209 AGGACCTGTGGCGGCTGGTCGAGACGGGCACGGACGCGATCGGCGGCTTCCCCACCGACC 41268

Qy        152 ---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGACGCCCG 208
Db      41269 GTGGCTGGGACGTCGACGCCGTGTACGACCCGGATCCTGAGTCGCGGAACACCACCTACT 41328

Qy        209 TTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCATCT 268
Db      41329 GCCGGGAGGGCGGGTTCCTGGCCGGGGCAGGAGACTTCGACGCCGCGTTCTTCGGGGTGT 41388

Qy        269 CGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCTGGG 328
Db      41389 CGCCGCACGAGGCCGTGGTCATGGACCCCCAGCAGCGGCTGCTTCTGGAGGTGTCCTGGG 41448

Qy        329 AGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAGTGT 388
Db      41449 AGGCGCTGGAGCGGTCCGGGACCGACCCGCACAGCCTGCGCGGCTCGCGCACCGGGGTCT 41508

Qy        389 TCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCGCAG 448
Db      41509 ACGTCGGTGCGGCCCACCAGGGGTACGCGGTCGACGCCGGTCAGGTGCCGGAGGGCGCGG 41568

Qy        449 AGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCTCGT 508
Db      41569 AG------GGGTTCCGGCTGACCGGCAGCGCCGACGCCGTCCTGTCCGGACGGATCTCGT 41622

Qy        509 ATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGCTGG 568
Db      41623 ACCTGCTCGGGCTGGAGGGTCCGGCCCTGACCGTCGAGACGGCCTGCTCGTCCTCGCTGG 41682

Qy        569 TGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGGCTG 628
Db      41683 TGGCGGTGCACCTCGCGGTGCAGGCGCTGCGCCGGGGCGAGTGCGGGCTGGCACTGGCCG 41742

Qy        629 GTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGGCGC 688
Db      41743 GCGGGGTCGCCGTGATGCCCGACCCGGCGGCATTCGTGGAGTTCTCCCGGCAGCGGGGCC 41802

Qy        689 TGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAGGCG 748
Db      41803 TCGCGGCGGACGGGCGCTGCCGGGCGTTCGGGGCGGGCGCGGACGGCACCGGCTGGGCGG 41862

Qy        749 AAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATCGGA 808
Db      41863 AGGGCGTCGGTGTGCTGGTCCTGCAACGGCTCTCCGACGCGGTGCGCGACGGCCGCTGGG 41922

Qy        809 TATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTGACCG 868
Db      41923 TGCTGGGCGTGATCCGGGGTTCGGCCGTCAACCAGGACGGGGCCAGCAACGGGCTGACCG 41982

Qy        869 TGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGCTGCG 928
Db      41983 CCCCGAGCGGCCCCGCCCAGCAGCGGGTCATCCGGCAGGCGCTGACCGACGCCCGGCTCG 42042
```

```
Qy      929 CCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGACCCCA 988
Db    42043 GCGCCGACCAGATCGACGCGGTCGAGGCGCACGGCACGGGCACCCGGCTCGGCGACCCGA 42102

Qy      989 TCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCGCTGC 1048
Db    42103 TCGAGGCGCAGGCGCTGATCGCCGCCTACGGCGCCGACCGGACCCCGGACCGGCCGCTCT 42162

Qy     1049 TGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACTGGGC 1108
Db    42163 GGCTCGGCTCGTTGAAGTCGAACATCGGGCACGCCCAGGCGGCGGCCGGCGTCGGCGGCC 42222

Qy     1109 TGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGCGCAGG 1168
Db    42223 TGATCAAGATGCTCCTGGCGATGCGGGCCGGGACGCTCCCACCCACCCTGCACGCCGACG 42282

Qy     1169 CGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCGGACAC 1228
Db    42283 TCCCGACCCCGCTGGTCGACTGGTCCGCCGGTGTCGTCCGGCTGTCGACCGGGGTGGTGC 42342

Qy     1229 CGTGGCCGGACTGG---AATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGAGCG 1285
Db    42343 CCTGGCCCGCGTTGCCCGGGGCGCCCCGCAGGGCCGGGATCTCCGCGTTCGGGGTGAGCG 42402

Qy     1286 GGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCGGCGC 1345
Db    42403 GCACCAACGCGCACGTGATCGTCGAGCAGCCGCCGCCGGTCCCGGTCGACGACCCGGCGC 42462

Qy     1346 C---------GGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCC 1396
Db    42463 CACCCACGAGGACCCTGCCGCTGGTGCCGTGGGTGCTCTCCGGCCGGACGGAGGCGGCGC 42522

Qy     1397 TGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGG 1456
Db    42523 TGCGCGCCCAGGCGGACCGGTTGCGTACGCACCTGGCGGCGCACCCCGACGCGGACCCGC 42582

Qy     1457 GCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGG 1516
Db    42583 TGGACGTGGGATTCTCCCTGGCCACCAGCCGGGCGCGCTGGAGCACCGGGCCGTGCTGG 42642

Qy     1517 CGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGT 1576
Db    42643 TGGCCGCCGACCGCGACGGCCTGCTCCGCCTCGTCGACGCGCTGGCCGCCGGCGAGCCGG 42702

Qy     1577 CGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAGCTCGCCTTTCTCTTCA 1636
Db    42703 CGGCGGGCCTGATCCGGGGCACGGTACGTCACGATCGCCG---GACCGGGTTCCTCTTCG 42759

Qy     1637 CCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGT 1696
Db    42760 CCGGGCAGGGCGGCCAGCGCGTCGGGATGGCGCGCGAACTGTACGAGGCGTTCCCCGCCT 42819

Qy     1697 TCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCC 1756
Db    42820 TCGCCGACGCCCTGGACCAGCTCGCCGCCCGGCTGGACCGGCACCTCGATCGTCCGCTGC 42879

Qy     1757 GCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCT 1816
Db    42880 TGCGGGTGCTGTTCGCCGAGCCGGGGTCGGACGACGCCCGGCTGCTCGACGGCACCCGGT 42939

Qy     1817 TCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGG 1876
Db    42940 ACGCGCAGGCCGCCCTCTTCGCCGTCGAGGTGGCGTTGTTCCGACTGGTCCACGGCTGGG 42999

Qy     1877 GTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGG 1936
Db    43000 GGGTCCGGCCCGACGTGCTGCTCGGCCACTCGGTGGGCGAGCTGGCGGCCGCGCACGTGG 43059

Qy     1937 CGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGC 1996
Db    43060 CCGGCGTACTCGACGTGGACGACGCGTGCGAGCTGGTCGCGGCGCGGGGCCGGCTGATGG 43119

Qy     1997 AGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTG 2056
Db    43120 GGGAGCTGCCGTCGGGCGGCGCGATGGTGGCGGTCCGGGCCACCGAGGAGGAGGTCGGGC 43179

Qy     2057 CTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCCGGACCAGG 2116
Db    43180 CCCTGCTCG------ACGGGCAGCGGGTCGCGGTGGCGGCGGTCAACGGCCCGCGCTCGG 43233

Qy     2117 TGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCG 2176
Db    43234 TCGTGGTCTCCGGCGACGAGGAGGCGGTGCTGGCCGTGGCCGCCCGGTGCGCCGCCCTCG 43293

Qy     2177 GGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGA 2236
Db    43294 GCCACCGGACGCGACGCCTCAACGTCAGCCACGCGTTCCACTCCCCGCACGTGGAGGCGA 43353

Qy     2237 TGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCC 2296
Db    43354 TGCTGGAGCCGTTCCGGCGGGTGGCGCGGGGCCTGACGTACCATGCCCCGACGATCCCGG 43413

Qy     2297 TGGTCAGCAATCTGA---GCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATT 2353
Db    43414 TGGTGTCGAACGCGACGGGCCGGCTCGCCACCGCCGACGCGCTGCGCGACCCCGGTTACT 43473

Qy     2354 GGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGG 2413
Db    43474 GGGTCCGGCACGTCCGCCAGCCCGTCCGGTTCCGGGACGGGGTGCGGGCCGCCCGCGACC 43533

Qy     2414 CCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCTGGTGCCTG 2473
Db    43534 AGGGGGCCACCGCCTTCGTCGGGCTCGGCCCGGACGGGGTGCTGTGCGCGTTGGCCGAGG 43593

Qy     2474 CCTGCAT---GCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGACG 2530
Db    43594 AGTGCCTCGGGCCCACCGGCGACGTGCTGCTGCTGCCGGTGCTGCGCCCCGGTCGGCCGG 43653

Qy     2531 AGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCCT 2590
Db    43654 AGCCCGCCACCCTGCTGGCCGCCCTGGCCGGGGCGTACGCCGGCGGCGCGGAAATGGACT 43713

Qy     2591 GGGCCGGCCTCTTCCC---CTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGC 2647
Db    43714 GGTCCCGGGTGTTCGCGGGCACCGGCGCGCGCAGGGTCGAGCTGCCCACGTACGCCTTCC 43773

Qy     2648 AGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGACCGCCGTG 2707
Db    43774 AGCACCGGCGCTACTGGCTGGCGCCGGGCCCGCCGTCGGCCCGCCGCGACGACGCCTG-- 43831

Qy     2708 CTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCGGC 2767
Db    43832 --GCGGTACCGGATCGCCTGGCGGCCCCTGCCGACCGTGCCCGCCGCCGCCGGGACCGAG 43889

Qy     2768 GCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTCG-AGGCC 2826
Db    43890 ACGGTGGCCGGGGCGTGGTTGCTGGTGGTCCCCGCCCACGACGGCGTCGCGTCGCTCGCC 43949

Qy     2827 GCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCACCTCGTG 2886
Db    43950 GACGCCGCCGAGCGGGCCGTGCACCGGGGCGGGGCCACGGTCACCCGGCTGACGGTGGAC 44009

Qy     2887 CTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTCGACGCG 2946
Db    44010 GCCGCCGACGTGGACCGGGACACCCTCGCCGCCGTGCTGACCGAGGCCGCCGCCGACGCG 44069

Qy     2947 GCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGACCTGCCG 3006
Db    44070 GACGGC------------GGGCCGGACGGGGTGCTCTGCCTGCTGGGCCTCGACGACCGG 44117

Qy     3007 GGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTCGCCGTG 3066
Db    44118 GCACATCCCCGGTCCGCCTCGGTGCCCCGCGGGGTGCTGGCGACCCTGTCCCTCGCCCAG 44177

Qy     3067 GGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCGGGAGCG 3126
```

```
Db      44178 GCCCTGACCGACCTGGGGGCCTCCGCGCGGCTGTGGTGCGTGACCCGGGGGGCGGTCGCC 44237

Qy       3127 TTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCGCTCTCG 3186

Db      44238 GTGACGCCCGGCGAGTCCCCGTCGGTCGCCGGAGCCCAGTTGTGGGGCTTCGGCCGCGTG 44297

Qy       3187 GCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTCGACAGA 3246

Db      44298 GCCGCGCTCGAACTCCCCCGGTCCTGGGGCGGCCTGGTGGACCTGCCGGTCGACCCGGAC 44357

Qy       3247 ATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGGGTCGGT 3306

Db      44358 GACCGGGACTGGGACCTGCTGCGGCGCGCGCTGCGCGGCCCGGAGGACCAGGTCGCGGTC 44417

Qy       3307 CTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCCGGCACG 3366

Db      44418 CGGGGGGCGGTCGGGTACGCCCGGCGGCTGGTCCCCG-CGCCCGCGCCCCGGGCCGAGCG 44476

Qy       3367 CCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCCCGCTCG 3426

Db      44477 GGCCTGGCGTCCGCGCGGCACGGTCCTGGTGACCGGCGGTACGGGCGCGCTCGGCGCGCA 44536

Qy       3427 GACCGGTTCGTCGCCGACGTGC 3448

Db      44537 CACGGCCCGCTGGCTGGCGCGC 44558
```

RESULT 10
US-10-212-962-2
; Sequence 2, Application US/10212962
; Publication No. US20030114450A1
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Santi, Daniel
; APPLICANT: Myles, David
; APPLICANT: Zq, Tian
; APPLICANT: Hutchinson, Richard
; APPLICANT: Johnson, Robert
; APPLICANT: Zhou, Yi-Qing
; APPLICANT: Feng, Li
; TITLE OF INVENTION: BENZOQUINONE ANZAMYCINS
; FILE REFERENCE: 30062-20075.00
; CURRENT APPLICATION NUMBER: US/10/212,962
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/310,779
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/389,255
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/393,929
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: No. US20030114450A1 yet assigned
; PRIOR FILING DATE: 2002-07-12
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 6360
;   TYPE: DNA
;   ORGANISM: Streptomyces geldanus
US-10-212-962-2

```
  Query Match             18.0%;  Score 765.8;  DB 14;  Length 6360;
  Best Local Similarity   56.1%;  Pred. No. 3.2e-159;
  Matches 1706;  Conservative    0;  Mismatches 1237;  Indels   98;  Gaps   10;

Qy         24 CGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGAT 83

Db         93 CGAGTCCACCGAACCGGTGGCGATCGTGGCGATGGCGTGCCGGCTGCCGGGTGGGGTGAC 152

Qy         84 CGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCC 143

Db        153 CGACCCGGAGTCGCTGTGGGAGCTGGTGGACGAGGGGCGGGACGCGATCGGGCCGTTCCC 212

Qy        144 CGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAA 200

Db        213 CACGGATCGCGGCTGGGACCTGGAGACCCTGTTCGACTCCGATCCGGACGCCGTGGGCAA 272

Qy        201 GACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTT 260

Db        273 GTCCTACGTACGCGAGGCGGGGTTCCTGGAGGGGGCGGCGGATTCGACGCCGCCTTCTT 332

Qy        261 CGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGT 320

Db        333 CGGCATCTCGCCGCGCGAGGCCCTGTCGCTGGACCCGCAGCAGCGGCTGCTGCTGGAGAC 392

Qy        321 GTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAAC 380

Db        393 CGCGTGGGAGACCTTCGAGCGGGCGGGGATGGATCCGCGGTCGGTGGAGGGCCGGGACAT 452

Qy        381 GGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGC 440

Db        453 CGCGGTGTTCGCCGGGGGCAGCGGCCAGGGGTACGGCGGCGGTCCGGGTGA------GGC 506

Qy        441 GTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCG 500

Db        507 GCCCAAGGGCCTGGAGGGCTATCTGGGGGTCGGCGCTTCCGGCAGTGTCATCTCCGGGCG 566

Qy        501 AATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTC 560

Db        567 CGTGTCGTACACGCTCGGGCTGACCGGTCCCGCCGTGACCGTGGACACCGCCTGCTCGTC 626

Qy        561 CTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGC 620

Db        627 CTCGCTGGTGGCCGCCCATCTCGCCGTGCAGGCGCTGCGGTCCGGCGAATGTTCCATGGC 686

Qy        621 CCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGAC 680

Db        687 GCTGGCCGGTGGTGTCGCCGTGATGGGCCAGCCCACCGCCTTCGTCGAGTTCTCCCGGCA 746

Qy        681 CCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGG 740

Db        747 GCGTGGCCTGGCGCCCGACGGGCGCTGCAAGTCCTTCGGCGCGGGCGCCGACGGCACCAC 806

Qy        741 ACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGG 800

Db        807 CTGGTCCGAAGGTGTCGGGCTCGTTCTGCTGGAGCGGCTGTCGGACGCCCGCCGCAACGG 866

Qy        801 CGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGG 860

Db        867 CCACGAAGTGCTGGCCGTGATCCGGGGCACCGCGGTCAACCAGGACGGCGCCTCCAACGG 926

Qy        861 TCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGC 920

Db        927 ACTCACCGCGCCCAACGGCCCCTCCCAGGAGCGGGTGATCCGCCAGGCCCTGTCCAACGC 986

Qy        921 AGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGG 980

Db        987 CGGGCTGACGGTGGCCGACGTGGACGCCGTCGAGGCCCACGGCACCGGCACCGCCCTCGG 1046

Qy        981 TGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCAC 1040

Db       1047 CGACCCCATCGAAGCCCAGGCCGTTCTCGCCACCTACGGCCAAAGCCGCCCGGAGGGCCG 1106

Qy       1041 GCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGAT 1100

Db       1107 GCCGCTGTGGCTCGGCTCCCTCAAGTCCAACATCGGCCACGCGCAGGCCGCAGCGGGCAT 1166

Qy       1101 CACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCA 1160

Db       1167 CGCCAGTGTCATCAAGACCGTCATGGCCTTACGCCACGGCCGGTTGCCGAAGACCCTCCA 1226

Qy       1161 CGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGC 1220

Db       1227 CGCCGAACAGCCCACCTCCCAGGTGAACTGGACGTCGGGCGCGGTGTCCCTGCTCGCCGA 1286

Qy       1221 CCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCAT 1280
```

```
Db   1287 GGCGCGGGCGTGGCCGGAGACCGGACACGCCCGCCGCGCCGGGATCTCCTCCTTCGGCGT 1346

Qy   1281 GAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGAC------------ 1328
Db   1347 CAGCGGGACGAACGCACACGTCATCCTGGAACAGGCCCCTGAGGAAGCCGAGGCGACCGG 1406

Qy   1329 ----------------------------------------GTGCACACCGCCGGCGCCGGAGC 1351
Db   1407 GGAGAACACCGCCGATCAGGAACCGCCCGTACGCTCGGCGGAGTCCGCCGACCCCGGCCC 1466

Qy   1352 GACCGGCA-----------GAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGGA 1400
Db   1467 GGTCGCCACCGGCCACGTGGTGCCGTGGCTGCTCTCGGGCCATACGCAGGAGGCGCTGCG 1526

Qy   1401 TGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCGA 1460
Db   1527 TGCCCAGGCCGCCCGGCTGCTGACCCAGGTGCGCGAGACGCCCTCCGACAGTCCGCGGGA 1586

Qy   1461 TGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGGC 1520
Db   1587 CGTGGGCTGGTCACTGGCCACCACCCGGACCCGGCTGGACCACCGCGCGGTCGTACTGTG 1646

Qy   1521 GACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGCC 1580
Db   1647 CGCCGATGCCGAGCAGGCCGTCGCGGGGCTGGAGGCGGTGGCCTCGGGCACGTCCGCCCG 1706

Qy   1581 CGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAGCTCGCCTTTCTCTTCACCGG 1640
Db   1707 GTCGGCGGTCACCGGGTCCGTGGCCTC------CGGAAAGGTGGCGGTGCTGTTCACCGG 1760

Qy   1641 ACAGGGGGCGCAGACGCTGGGCATGGCCGTGGGCTGTACGATGTATGGTCCGCGTTCCG 1700
Db   1761 GCAGGGCAGCCAGCGGGCCGGAATGGGCCGCGAACTGCACGGCGCCCACCCGGTGTTCGC 1820

Qy   1701 CGAGGCGTTCGA------------CCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCG 1748
Db   1821 GCGGGCCTTCGACGCCGTGTGCGCCCAGTTCGGCGACCTGCGCGACGGGGACGACAAGGT 1880

Qy   1749 GCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCA 1808
Db   1881 CTCGCTGGCCGAGGTGATCTTCGCCGAGGAGGGGTCGGCGACGGCAGCGCTGCTGGACCG 1940

Qy   1809 GACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCG 1868
Db   1941 GACCGAGTTCACCCAGCCCGCGCTGTTCGCGCTGGAGGTGGCGCTGTTCCGGCTCGTGGA 2000

Qy   1869 GTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGC 1928
Db   2001 GTCGTGGGGAGTGCGCCCCGCGTATGTGCTGGGCCACTCGATCGGCGAAGTGGCGGCGGC 2060

Qy   1929 CTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCG 1988
Db   2061 CCATGTGGCCGGGGTCCTGTCCCTGCCGGACGCCTGCACATTGGTGCGGGCGCGCGGGCG 2120

Qy   1989 CCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGA 2048
Db   2121 GCTGATGCAGCAACTCACCGCGACCGGGGCGATGGTCGCGGTGGAGGCGGCCGAGGACGA 2180

Qy   2049 TGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCC 2108
Db   2181 GGTGGCGCCGCTGCTCGCGGGGAAGGAGCACAAGGTCTCCATCGCCGCGGTCAACGGCCC 2240

Qy   2109 GGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGC 2168
Db   2241 GGCCTCCGTGGTCGTCTCCGGTGACGAGGACGTGGTCACGGCGGTGGCGGAGACGCTGGC 2300

Qy   2169 CGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCAT 2228
Db   2301 GCGGCAGGGCCGCAAGACCAAGCGGCTCGTGGTCTCGCACGCCTTCCACTCCCCCCACAT 2360

Qy   2229 GGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTC 2288
Db   2361 GGACGGGATGCTGGACGCGTTCCGCGAGGTGGCGTCGCGGCTGGCCTACGCGCCACCCCG 2420

Qy   2289 GATCGTCCTGGTCAGCAATCTGA---GCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCC 2345
Db   2421 GATACCCGTGGTGTCGAACCTCACCGGCGCGGTCGCCGATCCCGAGGAGCTGTGCTCCCC 2480

Qy   2346 GGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCT 2405
Db   2481 CGAGTACTGGGTACGGCATGCACGTGGCGCGGTGCGGTTCCTCGACGGTGTCCGCACACT 2540

Qy   2406 GCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCT 2465
Db   2541 GGCCGACGAGGGCGTGCGCACCCATCTGGAACTCGGCCCGGATGGGGTGCTGACCGCGAT 2600

Qy   2466 GGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCG 2525
Db   2601 GGGGCAGGACTGTCTGCCCGAGGCGGACGCGGCGTTCGTGCCGTCCCTGCGTCCGGGCGT 2660

Qy   2526 TGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGT 2585
Db   2661 CCAGGAGCCGCACGCGGTGCTGGCCGGGCTCGCCGGCCTGTACGTACGGGGTGTGCGGGT 2720

Qy   2586 CTCCTGGGCCGGCCTCTTCCCC---TCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCC 2642
Db   2721 GGACTGGGACGCGATGTTCGCCGGGTCCGGCGCCCGGCCCGTCGCCCTTCCCACGTACGC 2780

Qy   2643 TTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGACCG 2702
Db   2781 CTTCCAGCACGAGCACTACTGGCTGGAGCGGGCCGCCGGCTCCG-GCGACGTGGGCGCGG 2839

Qy   2703 CCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGA 2762
Db   2840 TGGGGCTCGGCGAGGCGGGCCATCCGCTGCTGGGCGCGGTGGTGCAGCTCCCGGAGACGG 2899

Qy   2763 CCGGCGCAGCGCTCGGCTCGACCATCCG----CCGCCCGAGAGCGGACGCCGGGAGAAGG 2818
Db   2900 GCGGGGTGCAGCTCAGCGGGCGGCTGTCGGTACGGGCCCAGCCCTGGCTGGGCGAACACG 2959

Qy   2819 TCGAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATC 2878
Db   2960 TCATCTCCGGGGCGGTGCTGGTGCCCGGCACCGCCATGGTGGAACTGGCCGTCCGCGCCG 3019

Qy   2879 ACCTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCG 2938
Db   3020 GGGACGAGACCGGCACCCCGGTGCTGGAGGAGCTGGTGATCGGGCAGCCGATGGTGCTGC 3079

Qy   2939 TCGACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCG 2979
Db   3080 CCGGCGACACCGCCCTCAGTGTCCAGGTCGTCGTGGGCGCG 3120
```

RESULT 11
US-10-156-761-2882
; Sequence 2882, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2882
; LENGTH: 10839

```
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(10839)
US-10-156-761-2882

  Query Match              17.8%;  Score 761.2;  DB 14;  Length 10839;
  Best Local Similarity    57.6%;  Pred. No. 3.1e-158;
  Matches 1589;  Conservative    0;  Mismatches 1103;  Indels  66;  Gaps   10;

Qy         26 CAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCG 85
Db       4685 CCGACGACGACCCGATCGTGATCGTGGGCATGGCCTGCCGCTTCCCCGGCGGGGTGCGCG 4744

Qy         86 ATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCG 145
Db       4745 GTCCCGAGGACCTCTGGCGTCTGCTCGTCGACAGCCGCGACGAGATGACGGACTTCCCCG 4804

Qy        146 CCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGACGC 205
Db       4805 CCGACCGAGGCTGGCACGGCCTGGCCATGAACGCCTTCATCGAGGAGTCCGGCGGAGCCC 4864

Qy        206 CCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCA 265
Db       4865 G------GCAGGGTGCCTTCCTGGCCGAGGCCGGCGACTTCGACGCGGCGTTCTTCGGGA 4918

Qy        266 TCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCT 325
Db       4919 TCTCGCCGCGTGAGGCACTGGCGATGGATCCGCAGCAGCGGTTGTTGCTGGAGACCTCGT 4978

Qy        326 GGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAG 385
Db       4979 GGGAGGCGCTGGAACGCGCCGGGTTCGATCCGGTCGGCCTGCGCGGCAGCCGCACCGGCG 5038

Qy        386 TGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCG 445
Db       5039 TCTTCGTCGGCGGCACCCCGCAGGAGTACACGACCGTCCTG------ATGAACTCGTCCG 5092

Qy        446 CAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCT 505
Db       5093 AGGCAGGCGGCGGTTACGCCCTCACCGGGGCGTCGGGCAGCGTGATGTCGGGCCGGGTCG 5152

Qy        506 CGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGC 565
Db       5153 CCTATGCCCTGGGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCGTGCTCCTCCTCGC 5212

Qy        566 TGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGG 625
Db       5213 TGGTCAGCCTGCACCTGGCAGGGCAGGCGCTGCGGGGCGGCGAGTGCGACCTGGCGCTGG 5272

Qy        626 CTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGG 685
Db       5273 TCGGCGGTGTGACCGTCATGGCCACGCCGGGAGCCTTCGTCGAGTTCTCGCGGCAGGGCG 5332

Qy        686 CGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAG 745
Db       5333 GTCTCGCGGGAGACGGCCGGTGCAAGGCGTTCGCCGCCGGTGCCGACGGCACGGGCTGGG 5392

Qy        746 GCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATC 805
Db       5393 GCGAGGGTGTCGGGATGCTGGCCGTGCAGCGGCTTTCGGACGCGGTGCGGGACGGGCGTC 5452

Qy        806 GGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTGA 865
Db       5453 CGGTGCTGGCGGTGGTGCGGGGTTCGGCGGTGAACTCGGACGGTGCGTCGAACGGTCTGA 5512

Qy        866 CCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGCT 925
Db       5513 CGGCGCCGAACGGTCCGTCGCAGCAGCGGGTGATCCGGCAGGCGCTCGCTTCGGCCGGGT 5572

Qy        926 GCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGACC 985
Db       5573 TGTCGGCGGCGGACGTGGACATGGTGGAGGCGCACGGCACGGGGACGTCGCTCGGCGACC 5632

Qy        986 CCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCGC 1045
Db       5633 CGATCGAGGCACAGGCGTTGCTGGCGACGTACGGTCAGGACCGTCCGGCGGACCGTCCGC 5692

Qy       1046 TGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACTG 1105
Db       5693 TGTGGCTGGGTTCGGTGAAGTCGAACATCGGCCACACCCAGTACGCCGCCGGTGTCGCCG 5752

Qy       1106 GGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGCGC 1165
Db       5753 GTGTCATCAAGTCCGTACTGGCCCTGCGCAACGGCCTGTTGCCCAAGACCCTGCACGTGG 5812

Qy       1166 AGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCGGA 1225
Db       5813 ACGAGCCCGACCCCCGAGGTGGACTGGTCCGCGGGCGCGGTGGAGCTGCTGACGGAGGGGC 5872

Qy       1226 CACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGAGCG 1285
Db       5873 GGGAGTGGCCGGAGACCGACGGGCCGCGCCGGGCGGGCGTGTCCTCGTTCGGGATCAGCG 5932

Qy       1286 GGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCG---- 1341
Db       5933 GCACGAACGCGCACGTGATTCTGGAGCAGGCGCCGTCGGTCGAGGAACGGACGCCGGTGT 5992

Qy       1342 --GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGG 1399
Db       5993 CTGCGGTTGACGGGTTGCTGGTGCCGTGGGTGGTGTCGGCGCGGTCGGAGGAGGCGTTGC 6052

Qy       1400 ATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCG 1459
Db       6053 GGGCGCAGGCCCGACGGCTTGCCGACCACGTACGGGGGAGCGAT---CTGCGGCCGGCCG 6109

Qy       1460 ATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGG 1519
Db       6110 ACGTGGGTCTGTCGTTGGCGGTGACGCGTGCGGGTCTGGAGCACCGGGCTGTTCTGACGG 6169

Qy       1520 CGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGC 1579
Db       6170 GGCGGGATCGGGAGGACTTCCTGGTCCAGTTGGCCGCGTTGGCGGAGGGTGCGGGCGCGG 6229

Qy       1580 CCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAGCTCGCCTTTCTCTTCACCG 1639
Db       6230 CGGGGATGGTGCGGGGTGTGGCCGGGGAG------GGCGGGACGGCGTTCCTGTTCACCG 6283

Qy       1640 GACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGTTCC 1699
Db       6284 GTCAGGGTGCGCAGCGGCTTGGTATGGGGCGTGAGCTGTACGAGCGTTTCCCGGTGTTCG 6343

Qy       1700 GCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCCGCG 1759
Db       6344 CGGCTGCGTTCGACGGGGTGTGTGCGCAGTTGGATCTGTTGCTGGAGCGGCCGGTGAAGG 6403

Qy       1760 AGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCTTCA 1819
Db       6404 AGGTCGTGTTCACCGA--------------TGCGCCGGCGTTGGACCGGACGGTGTTCA 6448

Qy       1820 CCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGGGTG 1879
Db       6449 CGCAGGCCGGTTTGTTCGCGTTGGAGGTGGCGCTGTTCGAGCTGGTGGGTTCGTGGGGTG 6508

Qy       1880 TAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGGCGG 1939
Db       6509 TGCGGGCGGATGTGCTGCTGGGGCATTCGATCGGTGAGCTGGCGGCGGCGTATGTGGCGG 6568

Qy       1940 GCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGCAGG 1999
Db       6569 GGGTGTGGTCGCTGGCGGATGCGTGCCGGATCGTGGCGGCGCGGGGCCGGTTGATGCAGG 6628

Qy       2000 CGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTGCTG 2059
Db       6629 CGCTGCCGGAGGGCGGTGCGATGGTCGCTGTCGAGGCGGCCGAGGACGA----------- 6677
```

```
Qy       2060 CGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCCGGACCAGGTGG 2119

Db       6678 -GCTGCCGGTGCTTCCCGGGGGTGTGTCGGTGGCTGCGGTGAACGGGCCCCGTTCGCTGG 6736

Qy       2120 TCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCGGGG 2179

Db       6737 TGCTGTCCGGTGATGAGGAGCCGGTGACCGCGCTCGCCCAAGCGTTTGCTGAACAGGGCA 6796

Qy       2180 CGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGATGC 2239

Db       6797 GGCGGACCAAGCAGTTGGCGGTCAGTCACGCCTTCCACTCCGCGCGGATGGAGCCGATGC 6856

Qy       2240 TGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCCTGG 2299

Db       6857 TGGCGGAGTTCGCGGAGACGCTCGCCGCGGTGGAGTTCCGTCCGCCGCGTATCCCGGTGG 6916

Qy       2300 TCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATTGGGTGC 2359

Db       6917 TGTCGAATGTGACGGGTGGGATGGCGGACGCCGAGTTCACCACTCCCGCCTACTGGGTGC 6976

Qy       2360 GCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGGCCGGTG 2419

Db       6977 GGCATGTGCGGGAGGCCGTTCGTTTCGCCGATGGTGTGGCGACGGTGCTGGCGCGGGGTG 7036

Qy       2420 CGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT------GCTCGGCCTGGTGCCTG 2473

Db       7037 TGGACAGGTTCCTCGAGCTGGGTCCGCGGGGTGCGCTGACCGCGATGGCCGAGGAGACCC 7096

Qy       2474 CCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGACGAGC 2533

Db       7097 TCGACCACACCGGCGCCGACGCCGTGTGCGTGCCCGTACTGCACCCCGAACGCCCCGAAG 7156

Qy       2534 CGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCCTGGG 2593

Db       7157 CGGACTGCCTGACCCTCGCGCTCGGCCGGATCCACGCGAGCGGTGCCCCGGTCGACTGGT 7216

Qy       2594 CCGGCCTCTTCCC---CTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGCAGC 2650

Db       7217 CCGGGCTGTTCGCCGGCACCGGCGCCCGCACCGTCGGACTGCCCACCTACGCCTTCCAGC 7276

Qy       2651 GCGAGCGCTACTGGATCGACACG---AAAGCCGACGACGCGGCGCGTGGCGACCGCCGTG 2707

Db       7277 ACAAGCGCTACTGGCTGGACTCGCGCACGACCGGCAGCGGCGACCCCGCGAGCATCGGAC 7336

Qy       2708 CTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCG 2765

Db       7337 TGACGGCGACCGGACACCCGCTGCTCGGCGCCGGAGTGGCGCTGCCCGACTCGGACGG 7394
```

RESULT 12
US-10-156-761-15102
; Sequence 15102, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15102
; LENGTH: 125746
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15102

Query Match 17.8%; Score 761.2; DB 14; Length 125746;
Best Local Similarity 57.6%; Pred. No. 2.5e-158;
Matches 1589; Conservative 0; Mismatches 1103; Indels 66; Gaps 10;

```
Qy         26 CAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCG 85

Db      89070 CCGACGACGACCCGATCGTGATCGTGGGCATGGCCTGCCGCTTCCCCGGCGGGGTGCGCG 89129

Qy         86 ATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCG 145

Db      89130 GTCCCGAGGACCTCTGGCGTCTGCTCGTCGACAGCCGCGACGAGATGACGGACTTCCCCG 89189

Qy        146 CCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGACGC 205

Db      89190 CCGACCGAGGCTGGCACGGCCTGGCCATGAACGCCTTCATCGAGGAGTCCGGCGGAGCCC 89249

Qy        206 CCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCA 265

Db      89250 G------GCAGGGTGCCTTCCTGGCCGAGGCCGGCGACTTCGACGCGGCGTTCTTCGGGA 89303

Qy        266 TCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCT 325

Db      89304 TCTCGCCGCGTGAGGCACTGGCGATGGATCCGCAGCAGCGGTTGTTGCTGGAGACCTCGT 89363

Qy        326 GGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAG 385

Db      89364 GGGAGGCGCTGGAACGCGCCGGGTTCGATCCGGTCGGCCTGCGCGGCAGCCGCACCGGCG 89423

Qy        386 TGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCG 445

Db      89424 TCTTCGTCGGCGGCACCCCGCAGGAGTACACGACCGTCCTG------ATGAACTCGTCCG 89477

Qy        446 CAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCT 505

Db      89478 AGGCAGGCGGCGGTTACGCCCTCACCGGGGCGTCGGGCAGCGTGATGTCGGGCCGGGTCG 89537

Qy        506 CGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGC 565

Db      89538 CCTATGCCCTGGGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCGTGCTCCTCCTCGC 89597

Qy        566 TGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGG 625

Db      89598 TGGTCAGCCTGCACCTGGCAGGGCAGGCGCTGCGGGGCGGCGAGTGCGACCTGGCGCTGG 89657

Qy        626 CTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGG 685

Db      89658 TCGGCGGTGTGACCGTCATGGCCACGCCGGGAGCCTTCGTCGAGTTCTCGCGGCAGGGCG 89717

Qy        686 CGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAG 745

Db      89718 GTCTCGCGGGAGACGGCCGGTGCAAGGCGTTCGCCGCCGGTGCCGACGGCACGGGCTGGG 89777

Qy        746 GCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATC 805

Db      89778 GCGAGGGTGTCGGGATGCTGGCCGTGCAGCGGCTTTCGGACGCGGTGCGGGACGGGCGTC 89837

Qy        806 GGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTGA 865

Db      89838 CGGTGCTGGCGGTGGTGCGGGGTTCGGCGGTGAACTCGGACGGTGCGTCGAACGGTCTGA 89897

Qy        866 CCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGCT 925

Db      89898 CGGCGCCGAACGGTCCGTCGCAGCAGCGGGTGATCCGGCAGGCGCTCGCTTCGGCCGGGT 89957

Qy        926 GCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGACC 985

Db      89958 TGTCGGCGGCGGACGTGGACATGGTGGAGGCGCACGGCACGGGGACGTCGCTCGGCGACC 90017

Qy        986 CCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCGC 1045
```

```
Db     90018 CGATCGAGGCACAGGCGTTGCTGGCGACGTACGGTCAGGACCGTCCGGCGGACCGTCCGC 90077

Qy      1046 TGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACTG 1105
Db     90078 TGTGGCTGGGTTCGGTGAAGTCGAACATCGGCCACACCCAGTACGCCGCCGGTGTCGCCG 90137

Qy      1106 GGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGCGC 1165
Db     90138 GTGTCATCAAGTCCGTACTGGCCCTGCGCAACGGCCTGTTGCCCAAGACCCTGCACGTGG 90197

Qy      1166 AGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCGGA 1225
Db     90198 ACGAGCCGACCCCCGAGGTGGACTGGTCCGCGGGCGCGGTGGAGCTGCTGACGGAGGGGC 90257

Qy      1226 CACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGAGCG 1285
Db     90258 GGGAGTGGCCGGAGACCGACGGGCCGCGCCGGGCGGGCGTGTCCTCGTTCGGGATCAGCG 90317

Qy      1286 GGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCG---- 1341
Db     90318 GCACGAACGCGCACGTGATTCTGGAGCAGGCGCCGTCGGTCGAGGAACGGACGCCGGTGT 90377

Qy      1342 --GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGG 1399
Db     90378 CTGCGGTTGACGGGTTGCTGGTGCCGTGGGTGGTGTCGGCGCGGTCGGAGGAGGCGTTGC 90437

Qy      1400 ATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCG 1459
Db     90438 GGGCGCAGGCCCGACGGCTTGCCGACCACGTACGGGGGAGCGAT---CTGCGGCCGGCCG 90494

Qy      1460 ATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGG 1519
Db     90495 ACGTGGGTCTGTCGTTGGCGGTGACGCGTGCGGGTCTGGAGCACCGGGCTGTTCTGACGG 90554

Qy      1520 CGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGC 1579
Db     90555 GGCGGGATCGGGAGGACTTCCTGGTCCAGTTGGCCGCGTTGGCGGAGGGTGCGGGCGCGG 90614

Qy      1580 CCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAGCTCGCCTTTCTCTTCACCG 1639
Db     90615 CGGGGATGGTGCGGGGTGTGGCCGGGGAG------GGCGGGACGGCGTTCCTGTTCACCG 90668

Qy      1640 GACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGTTCC 1699
Db     90669 GTCAGGGTGCGCAGCGGCTTGGTATGGGGCGTGAGCTGTACGAGCGTTTCCCGGTGTTCG 90728

Qy      1700 GCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCCGCG 1759
Db     90729 CGGCTGCGTTCGACGGGTGTGTGCGCAGTTGGATCTGTTGCTGGAGCGGCCGGTGAAGG 90788

Qy      1760 AGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCTTCA 1819
Db     90789 AGGTCGTGTTCACCGA---------------TGCGCCGGCGTTGGACCGGACGGTGTTCA 90833

Qy      1820 CCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGGGTG 1879
Db     90834 CGCAGGCCGGTTTGTTCGCGTTGGAGGTGGCGCTGTTCGAGCTGGTGGGTTCGTGGGGTG 90893

Qy      1880 TAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGGCGG 1939
Db     90894 TGCGGGCGGATGTGCTGCTGGGGCATTCGATCGGTGAGCTGGCGGCGGCGTATGTGGCGG 90953

Qy      1940 GCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGCAGG 1999
Db     90954 GGGTGTGGTCGCTGGCGGATGCGTGCCGGATCGTGGCGGCGCGGGGCCGGTTGATGCAGG 91013

Qy      2000 CGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTGCTG 2059
Db     91014 CGCTGCCGGAGGGCGGTGCGATGGTCGCTGTCGAGGCGGCCGAGGACGA----------- 91062

Qy      2060 CGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCCGGACCAGGTGG 2119
Db     91063 -GCTGCCGGTGCTTCCCGGGGGTGTGTCGGTGGCTGCGGTGAACGGGCCCCGTTCGCTGG 91121

Qy      2120 TCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCGGGG 2179
Db     91122 TGCTGTCCGGTGATGAGGAGCCGGTGACCGCGCTCGCCCAAGCGTTTGCTGAACAGGGCA 91181

Qy      2180 CGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGATGC 2239
Db     91182 GGCGGACCAAGCAGTTGGCGGTCAGTCACGCCTTCCACTCCGCGCGGATGGAGCCGATGC 91241

Qy      2240 TGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCCTGG 2299
Db     91242 TGGCGGAGTTCGCGGAGACGCTCGCCGCGGTGGAGTTCCGTCCGCCGCGTATCCCGGTGG 91301

Qy      2300 TCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATTGGGTGC 2359
Db     91302 TGTCGAATGTGACGGGTGGGATGGCGGACGCCGAGTTCACCACTCCCGCCTACTGGGTGC 91361

Qy      2360 GCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGGCCGGTG 2419
Db     91362 GGCATGTGCGGGAGGCCGTTCGTTTCGCCGATGGTGTGGCGACGGTGCTGGCGCGGGGTG 91421

Qy      2420 CGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT------GCTCGGCCTGGTGCCTG 2473
Db     91422 TGGACAGGTTCCTCGAGCTGGGTCCGCGGGGTGCGCTGACCGCGATGGCCGAGGAGACCC 91481

Qy      2474 CCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGACGAGC 2533
Db     91482 TCGACCACACCGGCGCCGACGCCGTGTGCGTGCCCGTACTGCACCCCGAACGCCCCGAAG 91541

Qy      2534 CGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCCTGGG 2593
Db     91542 CGGACTGCCTGACCCTCGCGCTCGGCCGGATCCACGCGAGCGGTGCCCCGGTCGACTGGT 91601

Qy      2594 CCGGCCTCTTCCC---CTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGCAGC 2650
Db     91602 CCGGGCTGTTCGCCGGCACCGGCGCCCGCACCGTCGGACTGCCCACCTACGCCTTCCAGC 91661

Qy      2651 GCGAGCGCTACTGGATCGACACG---AAAGCCGACGACGCGGCGCGTGGCGACCGCCGTG 2707
Db     91662 ACAAGCGCTACTGGCTGGACTCGCGCACGACCGGCAGCGGCGACCCCGCGAGCATCGGAC 91721

Qy      2708 CTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCG 2765
Db     91722 TGACGGCGACCGGACACCCGCTGCTCGGCGCCGGAGTGGCGCTGCCCGACTCGGACGG 91779
```

RESULT 13
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:

```
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

  Query Match             17.8%;  Score 761.2;  DB 14;  Length 9025608;
  Best Local Similarity   57.6%;  Pred. No. 1.8e-158;
  Matches 1589;  Conservative    0;  Mismatches 1103;  Indels   66;  Gaps   10;

Qy         26 CAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCG 85
Db    3562589 CCGACGACGACCCGATCGTGATCGTGGGCATGGCCTGCCGCTTCCCCGGCGGGGTGCGCG 3562530

Qy         86 ATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCG 145
Db    3562529 GTCCCGAGGACCTCTGGCGTCTGCTCGTCGACAGCCGCGACGAGATGACGGACTTCCCCG 3562470

Qy        146 CCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGACGC 205
Db    3562469 CCGACCGAGGCTGGCACGGCCTGGCCATGAACGCCTTCATCGAGGAGTCCGGCGGAGCCC 3562410

Qy        206 CCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCA 265
Db    3562409 G------GCAGGGTGCCTTCCTGGCCGAGGCCGGCGACTTCGACGCGGCGTTCTTCGGGA 3562356

Qy        266 TCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCT 325
Db    3562355 TCTCGCCGCGTGAGGCACTGGCGATGGATCCGCAGCAGCGGTTGTTGCTGGAGACCTCGT 3562296

Qy        326 GGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAG 385
Db    3562295 GGGAGGCGCTGGAACGCGCCGGGTTCGATCCGGTCGGCCTGCGCGGCAGCCGCACCGGCG 3562236

Qy        386 TGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCG 445
Db    3562235 TCTTCGTCGGCGGCACCCCGCAGGAGTACACGACCGTCCTG------ATGAACTCGTCCG 3562182

Qy        446 CAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCT 505
Db    3562181 AGGCAGGCGGCGGTTACGCCCTCACCGGGGCGTCGGGCAGCGTGATGTCGGGCCGGGTCG 3562122

Qy        506 CGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGC 565
Db    3562121 CCTATGCCCTGGGCCTGGAGGGCCCGGCCGTCACGGTCGACACCGCGTGCTCCTCCTCGC 3562062

Qy        566 TGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGG 625
Db    3562061 TGGTCAGCCTGCACCTGGCAGGGCAGGCGCTGCGGGGCGGCGAGTGCGACCTGGCGCTGG 3562002

Qy        626 CTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGG 685
Db    3562001 TCGGCGGTGTGACCGTCATGGCCACGCCGGGAGCCTTCGTCGAGTTCTCGCGGCAGGGCG 3561942

Qy        686 CGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAG 745
Db    3561941 GTCTCGCGGGAGACGGCCGGTGCAAGGCGTTCGCCGCCGGTGCCGACGGCACGGGCTGGG 3561882

Qy        746 GCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATC 805
Db    3561881 GCGAGGGTGTCGGGATGCTGGCCGTGCAGCGGCTTTCGGACGCGGTGCGGGACGGCGTC 3561822

Qy        806 GGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTGA 865
Db    3561821 CGGTGCTGGCGGTGGTGCGGGGTTCGGCGGTGAACTCGGACGGTGCGTCGAACGGTCTGA 3561762

Qy        866 CCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGCT 925
Db    3561761 CGGCGCCGAACGGTCCGTCGCAGCAGCGGGTGATCCGGCAGGCGCTCGCTTCGGCCGGGT 3561702

Qy        926 GCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGACC 985
Db    3561701 TGTCGGCGGCGGACGTGGACATGGTGGAGGCGCACGGCACGGGGACGTCGCTCGGCGACC 3561642

Qy        986 CCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCGC 1045
Db    3561641 CGATCGAGGCACAGGCGTTGCTGGCGACGTACGGTCAGGACCGTCCGGCGGACCGTCCGC 3561582

Qy       1046 TGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACTG 1105
Db    3561581 TGTGGCTGGGTTCGGTGAAGTCGAACATCGGCCACACCCAGTACGCCGCCGGTGTCGCCG 3561522

Qy       1106 GGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGCGC 1165
Db    3561521 GTGTCATCAAGTCCGTACTGGCCCTGCGCAACGGCCTGTTGCCCAAGACCCTGCACGTGG 3561462

Qy       1166 AGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCGGA 1225
Db    3561461 ACGAGCCGACCCCCGAGGTGGACTGGTCCGCGGGCGCGGTGGAGCTGCTGACGGAGGGGC 3561402

Qy       1226 CACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGTGAGCTCGTTCGGCATGAGCG 1285
Db    3561401 GGGAGTGGCCGGAGACCGACGGGCCGCGCCGGGCGGGCGTGTCCTCGTTCGGGATCAGCG 3561342

Qy       1286 GGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCG---- 1341
Db    3561341 GCACGAACGCGCACGTGATTCTGGAGCAGGCGCCGTCGGTCGAGGAACGGACGCCGGTGT 3561282

Qy       1342 --GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGG 1399
Db    3561281 CTGCGGTTGACGGGTTGCTGGTGCCGTGGGTGGTGTCGGCGCGGTCGGAGGAGGCGTTGC 3561222

Qy       1400 ATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCG 1459
Db    3561221 GGGCGCAGGCCCGACGGCTTGCCGACCACGTACGGGGGAGCGAT---CTGCGGCCGGCCG 3561165

Qy       1460 ATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGG 1519
Db    3561164 ACGTGGGTCTGTCGTTGGCGGTGACGCGTGCGGGTCTGGAGCACCGGGCTGTTCTGACGG 3561105

Qy       1520 CGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGC 1579
Db    3561104 GGCGGGATCGGGAGGACTTCCTGGTCCAGTTGGCCGCGTTGGCGGAGGGTGCGGGCGCGG 3561045

Qy       1580 CCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAGCTCGCCTTTCTCTTCACCG 1639
Db    3561044 CGGGGATGGTGCGGGGTGTGGCCGGGGAG------GGCGGGACGGCGTTCCTGTTCACCG 3560991

Qy       1640 GACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGTTCC 1699
Db    3560990 GTCAGGGTGCGCAGCGGCTTGGTATGGGGCGTGAGCTGTACGAGCGTTTCCCGGTGTTCG 3560931

Qy       1700 GCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCCGCG 1759
Db    3560930 CGGCTGCGTTCGACGGGGTGTGTGCGCAGTTGGATCTGTTGCTGGAGCGGCCGGTGAAGG 3560871

Qy       1760 AGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCTTCA 1819
Db    3560870 AGGTCGTGTTCACCGA--------------TGCGCCGGCGTTGGACCGGACGGTGTTCA 3560826

Qy       1820 CCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGGGTG 1879
Db    3560825 CGCAGGCCGGTTTGTTCGCGTTGGAGGTGGCGCTGTTCGAGCTGGTGGGTTCGTGGGGTG 3560766

Qy       1880 TAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGGCGG 1939
Db    3560765 TGCGGGCGGATGTGCTGCTGGGGCATTCGATCGGTGAGCTGGCGGCGGCGTATGTGGCGG 3560706

Qy       1940 GCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGCAGG 1999
Db    3560705 GGGTGTGGTCGCTGGCGGATGCGTGCCGGATCGTGGCGGCGCGGGGCCGGTTGATGCAGG 3560646

Qy       2000 CGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTGCTG 2059
Db    3560645 CGCTGCCCGGAGGGCGGTGCGATGGTCGCTGTCGAGGCGGCCGAGGACGA----------- 3560597

Qy       2060 CGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCCGGACCAGGTGG 2119
```

```
Db    3560596 -GCTGCCGGTGCTTCCCGGGGGTGTGTCGGTGGCTGCGGTGAACGGGCCCCGTTCGCTGG 3560538

Qy       2120 TCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCGGGG 2179

Db    3560537 TGCTGTCCGGTGATGAGGAGCCGGTGACCGCGCTCGCCCAAGCGTTTGCTGAACAGGGCA 3560478

Qy       2180 CGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGATGC 2239

Db    3560477 GGCGGACCAAGCAGTTGGCGGTCAGTCACGCCTTCCACTCCGCGCGGATGGAGCCGATGC 3560418

Qy       2240 TGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCCTGG 2299

Db    3560417 TGGCGGAGTTCGCGGAGACGCTCGCCGCGGTGGAGTTCCGTCCGCCGCGTATCCCGGTGG 3560358

Qy       2300 TCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATTGGGTGC 2359

Db    3560357 TGTCGAATGTGACGGGTGGGATGGCGGACGCCGAGTTCACCACTCCCGCCTACTGGGTGC 3560298

Qy       2360 GCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGGCCGGTG 2419

Db    3560297 GGCATGTGCGGGAGGCCGTTCGTTTCGCCGATGGTGTGGCGACGGTGCTGGCGCGGGGTG 3560238

Qy       2420 CGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT------GCTCGGCCTGGTGCCTG 2473

Db    3560237 TGGACAGGTTCCTCGAGCTGGGTCCGCGGGGTGCGCTGACCGCGATGGCCGAGGAGACCC 3560178

Qy       2474 CCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGACGAGC 2533

Db    3560177 TCGACCACACCGGCGCCGACGCCGTGTGCGTGCCCGTACTGCACCCCGAACGCCCCGAAG 3560118

Qy       2534 CGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCCTGGG 2593

Db    3560117 CGGACTGCCTGACCCTCGCGCTCGGCCGGATCCACGCGAGCGGTGCCCCGGTCGACTGGT 3560058

Qy       2594 CCGGCCTCTTCCC---CTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGCAGC 2650

Db    3560057 CCGGGCTGTTCGCCGGCACCGGCGCCCGCACCGTCGGACTGCCCACCTACGCCTTCCAGC 3559998

Qy       2651 GCGAGCGCTACTGGATCGACACG---AAAGCCGACGACGCGGCGCGTGGCGACCGCCGTG 2707

Db    3559997 ACAAGCGCTACTGGCTGGACTCGCGCACGACCGGCAGCGGCGACCCCGCGAGCATCGGAC 3559938

Qy       2708 CTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCG 2765

Db    3559937 TGACGGCGACCGGACACCCGCTGCTCGGCGCCGGAGTGGCGCTGCCCGACTCGGACGG 3559880
```

RESULT 14
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

```
  Query Match             17.4%;  Score 742;  DB 10;  Length 15872;
  Best Local Similarity   55.7%;  Pred. No. 5e-154;
  Matches 1594;  Conservative    0;  Mismatches 1215;  Indels   53;  Gaps    7;

Qy         26 CAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCG 85

Db       8490 CCGACGACGATCCGATCGCGATCGTGGGCATGGCATGCCGCTATCCGGGTGGTGTGTCGT 8549

Qy         86 ATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCG 145

Db       8550 CGCCGGAGGACCTGTGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCG 8609

Qy        146 CCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGA 202

Db       8610 TCAACCGCGGCTGGGACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCA 8669

Qy        203 CGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCG 262

Db       8670 CGTACTGCCGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCG 8729

Qy        263 GCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGT 322

Db       8730 GCATCTCGCCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGT 8789

Qy        323 GCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGG 382

Db       8790 CCTGGGAGGCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTG 8849

Qy        383 GAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGT 442

Db       8850 GTGTCTACGTGGGCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGG 8909

Qy        443 CCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAA 502

Db       8910 GCTCGGAG------GGCTATCTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCA 8963

Qy        503 TCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCT 562

Db       8964 TCTCCTACGCGCTCGGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCT 9023

Qy        563 CGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCC 622

Db       9024 CGCTGGTGGCGCTGCATCTGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGC 9083

Qy        623 TGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCC 682

Db       9084 TGGCGGGCGGGGTGGCGGTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGA 9143

Qy        683 GGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGAC 742

Db       9144 AGGGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCT 9203

Qy        743 GAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCG 802

Db       9204 GGGCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGC 9263

Qy        803 ATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTC 862

Db       9264 ACACGGTCCTCGGCCTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGC 9323

Qy        863 TGACCGTGCCGAACGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAG 922

Db       9324 TGACCGCGCCCAACGGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCG 9383

Qy        923 GCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTG 982

Db       9384 GGCTGTCCCCGGAGGACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCG 9443

Qy        983 ACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGC 1042

Db       9444 ACCCCATCGAGGCCGGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACC 9503

Qy       1043 CGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCA 1102

Db       9504 CGCTGTGGCTCGGCTCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCG 9563
```

```
Qy       1103 CTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACG 1162
Db       9564 GCGGCGTCATCAAGATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACG 9623
Qy       1163 CGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCC 1222
Db       9624 CCGACGAGCCGACCCCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCG 9683
Qy       1223 GGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGA 1282
Db       9684 AGGTGCCGTGGCAGCGGACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCG 9743
Qy       1283 GCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCGG 1342
Db       9744 GCGGCACCAATGCCCATGTCGTCCTCGAAGAGGCACCCGCCCCGCCCGCGCGCGGAACCGG 9803
Qy       1343 CGCCGGAGCGACC---------------------------------GGCAGAGCTGCTGG 1369
Db       9804 CCGGGGAGGCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGGCCCCTGGCCTGGG 9863
Qy       1370 TGCTGTCGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATC 1429
Db       9864 TGGTCTCCGGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCACC 9923
Qy       1430 TGGAGACCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCA 1489
Db       9924 TCTCCCGCACCCCCGGGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCGCCACGCGCG 9983
Qy       1490 GCGCGATGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCC 1549
Db       9984 CAGCCTTTGACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGCCC 10043
Qy       1550 TGGACGCTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCT 1609
Db      10044 TGGACGCGTTGGCCGAAGGACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGA---CC 10100
Qy       1610 CACGCGGCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCC 1669
Db      10101 GGGACGGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCCGGGATGGCCC 10160
Qy       1670 GTGGGCTGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGT 1729
Db      10161 ACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTGACGGACCGTC 10220
Qy       1730 TCAACCAGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCG 1789
Db      10221 TCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGCTGGACGCCCGACCCGGCTCGCCCG 10280
Qy       1790 ACGCCGCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATG 1849
Db      10281 AAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCGGCGCTCTTCGCCGTCGAGGTGG 10340
Qy       1850 CGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCA 1909
Db      10341 CGCTCCACCGGCTGCTGGAGCACTGGGGATGCGCCCCGACCTGCTGCTGGGGCACTCGG 10400
Qy       1910 TCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCC 1969
Db      10401 TGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTGCTCGATCTCGACGACGCCTGCGCGC 10460
Qy       1970 TGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGA 2029
Db      10461 TGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTGCCGCCCGGCGGCGCGATGGTCTCCG 10520
Qy       2030 TCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGA 2089
Db      10521 TGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTGGCCGGCCGCGAGGACGCCGTCTGCG 10580
Qy       2090 TCGCCGCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATG 2149
Db      10581 TCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCG 10640
Qy       2150 CGATCGCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATG 2209
Db      10641 AGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCACG 10700
Qy       2210 CGTTCCACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGG 2269
Db      10701 CCTTCCACTCACCCCTGATGGACGGCATGCTCGCCGGATTCCGGGAGGTCGCCGCCGGCC 10760
Qy       2270 TGAGCTACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAG---GCTTGCA 2326
Db      10761 TGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCCACGGTCACGGGGCGGCCCGCCCGCC 10820
Qy       2327 CAGACGAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCG 2386
Db      10821 CCGGTGAAGTCACCGGCCCCGACTACTGGGTGGCCCAGGTCCGTGAGCCCGTGCGCTTCG 10880
Qy       2387 CGGATGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGA 2446
Db      10881 CGGACGCGGTCCGCACGGCACACCGCCTCGGAGCCCGCACCTTCCTGGAGACCGGCCCGG 10940
Qy       2447 AATCGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCG 2506
Db      10941 ACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTGGAGGACGACACCGTGGCCCTGCTGC 11000
Qy       2507 CATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCT 2566
Db      11001 ---CGGCGATCCACAAGCCCGGCACCGCGCCGCACGGTCCGGCGGCTCCCGGCGCGCTGC 11057
Qy       2567 GGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGC 2626
Db      11058 GGGCGGCCGCCGCCGCGTACGGCCGGGGCGCCCGGGTGGACT--GGGCCGGGATGCACGC 11115
Qy       2627 CGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACG 2686
Db      11116 CGACGGCCCCGAGGGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTTCCGGCACCG 11175
Qy       2687 CGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCG 2746
Db      11176 CCGCTACTGGCTCGCCCCGGGCCGCGCGGCGGACACCGACGACTGGATGTACCGGATCGG 11235
Qy       2747 CGGTGCGCGGCGGCGACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGAC 2806
Db      11236 CTGGGACCGGCTGCCGGCTGTGACCGGCGGGGCCCGGACCGCCGGCCGCTGGCTGGTGAT 11295
Qy       2807 GCCGGGAGAAGGTCGAGGCCGCCGGCGACCGTCCGTTCCGGC 2848
Db      11296 CCACCCCGACAGCCCGCGCTGCCGGGAGCTGTCCGGCCACGC 11337
```

RESULT 15
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-1

Query Match 17.4%; Score 742; DB 10; Length 15872;

```
Best Local Similarity   55.7%;  Pred. No. 5e-154;
Matches 1594;  Conservative    0;  Mismatches 1215;  Indels   53;  Gaps    7;

Qy       26 CAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCG 85
Db     8490 CCGACGACGATCCGATCGCGATCGTGGGCATGGCATGCCGCTATCCGGTGGTGTGTCGT 8549

Qy       86 ATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCG 145
Db     8550 CGCCGGAGGACCTGTGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCG 8609

Qy      146 CCGAAC---GCTGGGATGCAGCAGCGTGGTTGATCCCGACCCCGATGCCCCGGGGAAGA 202
Db     8610 TCAACCGCGGCTGGGACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCA 8669

Qy      203 CGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCG 262
Db     8670 CGTACTGCCGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCG 8729

Qy      263 GCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGT 322
Db     8730 GCATCTCGCCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGT 8789

Qy      323 GCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGG 382
Db     8790 CCTGGGAGGCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTG 8849

Qy      383 GAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGT 442
Db     8850 GTGTCTACGTGGGCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGG 8909

Qy      443 CCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAA 502
Db     8910 GCTCGGAG------GGCTATCTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCA 8963

Qy      503 TCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCT 562
Db     8964 TCTCCTACGCGCTCGGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCT 9023

Qy      563 CGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCC 622
Db     9024 CGCTGGTGGCGCTGCATCTGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGC 9083

Qy      623 TGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCC 682
Db     9084 TGGCGGGCGGGGTGGCGGTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGA 9143

Qy      683 GGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGAC 742
Db     9144 AGGGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCT 9203

Qy      743 GAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCG 802
Db     9204 GGGCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGC 9263

Qy      803 ATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTC 862
Db     9264 ACACGGTCCTCGGCCTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGC 9323

Qy      863 TGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAG 922
Db     9324 TGACCGCGCCCAACGGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCG 9383

Qy      923 GCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTG 982
Db     9384 GGCTGTCCCCGGAGGACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCG 9443

Qy      983 ACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGC 1042
Db     9444 ACCCCATCGAGGCCGGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACC 9503

Qy     1043 CGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCA 1102
Db     9504 CGCTGTGGCTCGGCTCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCG 9563

Qy     1103 CTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACG 1162
Db     9564 GCGGCGTCATCAAGATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACG 9623

Qy     1163 CGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCC 1222
Db     9624 CCGACGAGCCGACCCCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCG 9683

Qy     1223 GGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGA 1282
Db     9684 AGGTGCCGTGGCAGCGGACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCG 9743

Qy     1283 GCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCGG 1342
Db     9744 GCGGCACCAATGCCCATGTCGTCCTCGAAGAGGCACCCGCCCCGCCCGCGCCGGAACCGG 9803

Qy     1343 CGCCGGAGCGACC-------------------------------GGCAGAGCTGCTGG 1369
Db     9804 CCGGGGAGGCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGCCCCTGGCCTGGG 9863

Qy     1370 TGCTGTCGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATC 1429
Db     9864 TGGTCTCCGGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCACC 9923

Qy     1430 TGGAGACCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCA 1489
Db     9924 TCTCCCGCACCCCCGGGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCGCCACGCGCG 9983

Qy     1490 GCGCGATGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCC 1549
Db     9984 CAGCCTTTGACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGCCC 10043

Qy     1550 TGGACGCTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCT 1609
Db    10044 TGGACGCGTTGGCCGAAGGACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGA---CC 10100

Qy     1610 CACGCGGCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCC 1669
Db    10101 GGGACGGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCCGGGATGGCCC 10160

Qy     1670 GTGGGCTGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGT 1729
Db    10161 ACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTGACGGACCGTC 10220

Qy     1730 TCAACCAGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCG 1789
Db    10221 TCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGCTGGACGCCCGACCCGGCTCGCCCG 10280

Qy     1790 ACGCCGCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATG 1849
Db    10281 AAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCGGCGCTCTTCGCCGTCGAGGTGG 10340

Qy     1850 CGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCA 1909
Db    10341 CGCTCCACCGGCTGCTGGAGCACTGGGGATGCGCCCCGACCTGCTGCTGGGGCACTCGG 10400

Qy     1910 TCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCC 1969
Db    10401 TGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTGCTCGATCTCGACGACGCCTGCGCGC 10460

Qy     1970 TGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGA 2029
Db    10461 TGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTGCCGCCCGGCGGCGCGATGGTCTCCG 10520

Qy     2030 TCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGA 2089
Db    10521 TGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTGGCCGGCCGCGAGGACGCCGTCTGCG 10580

Qy     2090 TCGCCGCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATG 2149
Db    10581 TCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCG 10640
```

```
Qy      2150 CGATCGCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATG 2209
Db     10641 AGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCACG 10700

Qy      2210 CGTTCCACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGG 2269
Db     10701 CCTTCCACTCACCCCTGATGGACGGCATGCTCGCCGGATTCCGGGAGGTCGCCGCCGGCC 10760

Qy      2270 TGAGCTACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAG---GCTTGCA 2326
Db     10761 TGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCCACGGTCACGGGGCGGCCCGCCCGCC 10820

Qy      2327 CAGACGAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCG 2386
Db     10821 CCGGTGAACTCACCGGCCCCGACTACTGGGTGGCCCAGGTCCGTGAGCCCGTGCGCTTCG 10880

Qy      2387 CGGATGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGA 2446
Db     10881 CGGACGCGGTCCGCACGGCACACCGCCTCGGAGCCCGCACCTTCCTGGAGACCGGCCCGG 10940

Qy      2447 AATCGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCG 2506
Db     10941 ACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTGGAGGACGACACCGTGGCCCTGCTGC 11000

Qy      2507 CATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCT 2566
Db     11001 ---CGGCGATCCACAAGCCCGGCACCGCGCCGCACGGTCCGGCGGCTCCCGGCGCGCTGC 11057

Qy      2567 GGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGC 2626
Db     11058 GGGCGGCCGCCGCCGCGTACGGCCGGGGCGCCCGGGTGGACT--GGGCCGGGATGCACGC 11115

Qy      2627 CGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACG 2686
Db     11116 CGACGGCCCCGAGGGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTTCCGGCACCG 11175

Qy      2687 CGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCG 2746
Db     11176 CCGCTACTGGCTCGCCCCGGGCCGCGCGGCGGACACCGACGACTGGATGTACCGGATCGG 11235

Qy      2747 CGGTGCGCGGCGGCGACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGAC 2806
Db     11236 CTGGGACCGGCTGCCGGCTGTGACCGGCGGGGCCCGGACCGCCGGCCGCTGGCTGGTGAT 11295

Qy      2807 GCCGGGAGAAGGTCGAGGCCGCCGGCGACCGTCCGTTCCGGC 2848
Db     11296 CCACCCCGACAGCCCGCGCTGCCGGGAGCTGTCCGGCCACGC 11337
```

Search completed: October 4, 2003, 12:39:42
Job time : 1051.7 secs

```
OM nucleic - nucleic search, using sw model

Run on:          October  3, 2003, 11:46:26 ; Search time 225.999 Seconds
                                              (without alignments)
                                              8331.640 Million cell updates/sec

Title:           US-10-014-717-1_COPY_7610_11875
Perfect score:   4266
Sequence:        1 gtggcggatcgtcccatcga..........gggaaatcattgccctatga 4266

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters:      1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                 1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                 2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                 3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                 4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                 5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                 6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4266	100.0	68750	3	US-09-335-409-1	Sequence 1, Appli
	2	4266	100.0	68750	4	US-09-568-102-1	Sequence 1, Appli
	3	4266	100.0	68750	4	US-09-567-969-1	Sequence 1, Appli
	4	4266	100.0	68750	4	US-09-568-480-1	Sequence 1, Appli
	5	4266	100.0	68750	4	US-09-568-486-1	Sequence 1, Appli
	6	4266	100.0	68750	4	US-09-568-472-1	Sequence 1, Appli
	7	4266	100.0	68750	4	US-09-567-899-1	Sequence 1, Appli
	8	4187.6	98.2	71989	4	US-09-443-501A-2	Sequence 2, Appli
	9	814.2	19.1	3978	4	US-09-266-965-19	Sequence 19, Appl
	10	814.2	19.1	12249	4	US-09-266-965-74	Sequence 74, Appl
	11	814.2	19.1	18331	4	US-09-266-965-96	Sequence 96, Appl
	12	803.2	18.8	33529	3	US-09-144-085-3	Sequence 3, Appli
	13	755.2	17.7	43280	2	US-08-804-227C-1	Sequence 1, Appli
	14	742	17.4	15872	3	US-09-105-537-1	Sequence 1, Appli
c	15	740.8	17.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	16	732	17.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	17	725	17.0	49377	1	US-08-764-233A-1	Sequence 1, Appli
	18	712.4	16.7	32155	4	US-08-311-731A-1	Sequence 1, Appli
	19	701.2	16.4	80161	3	US-09-036-987A-1	Sequence 1, Appli
	20	701.2	16.4	80161	3	US-09-370-700-1	Sequence 1, Appli
	21	701.2	16.4	80161	4	US-09-603-207-1	Sequence 1, Appli
	22	675.2	15.8	44377	2	US-08-804-227C-7	Sequence 7, Appli
	23	675.2	15.8	44377	2	US-08-804-198-1	Sequence 1, Appli
	24	671.8	15.7	28958	1	US-08-258-261B-6	Sequence 6, Appli
	25	671.8	15.7	28958	1	US-08-456-837-6	Sequence 6, Appli
	26	671.8	15.7	28958	1	US-08-457-342-6	Sequence 6, Appli
	27	671.8	15.7	28958	1	US-08-457-646A-6	Sequence 6, Appli
	28	671.8	15.7	28958	1	US-08-458-076A-6	Sequence 6, Appli
	29	671.8	15.7	28958	1	US-08-764-233A-4	Sequence 4, Appli
	30	671.8	15.7	28958	1	US-08-457-335A-6	Sequence 6, Appli
	31	671.8	15.7	28958	1	US-08-729-214-6	Sequence 6, Appli
	32	671.8	15.7	28958	3	US-09-028-934-6	Sequence 6, Appli
	33	663.2	15.5	20235	1	US-07-642-734C-3	Sequence 3, Appli
	34	663.2	15.5	20235	3	US-08-439-009A-3	Sequence 3, Appli
	35	662.2	15.5	13987	2	US-08-804-227C-13	Sequence 13, Appl
	36	649.4	15.2	47981	4	US-09-679-279-1	Sequence 1, Appli
	37	633	14.8	77536	4	US-09-410-551B-1	Sequence 1, Appli
	38	630.8	14.8	11219	1	US-07-642-734C-1	Sequence 1, Appli
	39	630.8	14.8	11219	3	US-08-439-009A-1	Sequence 1, Appli
	40	628.8	14.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	41	628.8	14.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	42	627.6	14.7	13842	3	US-09-105-537-30	Sequence 30, Appl
	43	627.6	14.7	36778	3	US-09-105-537-5	Sequence 5, Appli
	44	627.6	14.7	38506	3	US-09-320-878-19	Sequence 19, Appl
	45	627.6	14.7	38506	4	US-09-141-908-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-335-409-1

  Query Match             100.0%;  Score 4266;  DB 3;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969
```

```
Qy      361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy      421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy      481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy      541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy      601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy      661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy      721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy      781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy      841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy      901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy      961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy     1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy     1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy     1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy     1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy     1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy     1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy     1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy     1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy     1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy     1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy     1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy     1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy     1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy     1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy     1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy     1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy     1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy     2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy     2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy     2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy     2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy     2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy     2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy     2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy     2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy     2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875
```

```
RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,102
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             100.0%;  Score 4266;  DB 4;  Length 68750;
```

```
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy       1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy       1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy       1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy       2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy       2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769
```

```
Qy       2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy       2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy       2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy       2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875
```

```
RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             100.0%;  Score 4266;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy       1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy       1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy       1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy       2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy       2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy       2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy       2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy       2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy       2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569
```

```
Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875
```

RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

```
  Query Match             100.0%;  Score 4266;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989
```

```
Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy       1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy       1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy       1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy       2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy       2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy       2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy       2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy       2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy       2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy         3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy         3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy         3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy         3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy         3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 3900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 11509

Qy         3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy         3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy         4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy         4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy         4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy         4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy         4261 CTATGA 4266
                ||||||
Db        11870 CTATGA 11875
```

RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

```
  Query Match             100.0%;  Score 4266;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy            1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy           61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy          121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy          181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy          241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy          301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy          361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy          421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy          481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy          541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy          601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy          661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy          721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy          781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy          841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy          901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy          961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy         1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy         1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy         1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy         1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy         1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy         1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy         1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy         1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy         1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy         1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy         1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy         1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy         1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy         1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy         1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy         1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy         1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy         2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy         2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy         2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy         2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy         2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy         2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy         2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy         2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy         2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy         2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy         2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy         2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy         2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy         2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy         2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy         2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy         3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy         3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy         3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789
```

```
Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875


RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             100.0%;  Score 4266;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209
```

```
Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy       1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy       1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy       1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy       2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy       2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy       2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy       2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy       2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy       2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy         2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy         2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy         2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy         3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy         3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy         3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy         3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy         3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy         3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy         3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy         3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy         3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy         3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy         3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy         3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy         3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy         3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy         3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 11509

Qy         3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy         3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy         4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy         4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy         4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy         4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy         4261 CTATGA 4266
                ||||||
Db        11870 CTATGA 11875
```

RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

```
  Query Match             100.0%;  Score 4266;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy            1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy           61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy          121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy          181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy         241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy         301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy         361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy         421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy         481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy         541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy         601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy         661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy         721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy         781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy         841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy         901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy         961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy        1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy        1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy        1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy        1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy        1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy        1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy        1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy        1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy        1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy        1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy        1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy        1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy        1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy        1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy        1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy        1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy        1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy        2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy        2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy        2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy        2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy        2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy        2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009
```

```
Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875
```

RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19

```
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

  Query Match             98.2%;  Score 4187.6;  DB 4;  Length 71989;
  Best Local Similarity   98.9%;  Pred. No. 0;
  Matches 4217;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| | 
Db       1998 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGGGC 2057

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 2117

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db       2118 CGCGACACCGTCGGGCAAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 2177

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 GACCTCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 2237

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2238 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 2297

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2298 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 2357

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2358 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 2417

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2418 GCGCTGCCGCGAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 2477

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       2478 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGTCCTCGGGCTGCGAGGGCCGTGTGTCGCG 2537

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       2538 GTGGATACGGCCTATTCGTCCTCGCTCGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 2597

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2598 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 2657

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||| ||||||||| ||||||||||||||||||| ||||||
Db       2658 CTCGTGTGGCTCTCGAAGACCCGCGCGCTGGCCACGGACGGTCGCTGCAAGGCGTTTTCG 2717

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2718 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 2777

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2778 AGTGGAGCCCGCGCGGACGGCGACCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 2837

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGAGCTCCCAAGAAATCGTGCTG 900
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2838 CACGACGGAGCGAGCAGCGGTCTGACCGTGCCGAACGGAGCTCCCAAGAAATCGTGCTG 2897

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2898 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 2957

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2958 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 3017

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       3018 CTCGGGCGAGACGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 3077

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3078 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 3137

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3138 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 3197

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3198 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 3257

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3258 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 3317

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db       3318 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGGCCGGCAGAGCTGCTGGTGCTGTCGGCA 3377

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||| ||||| |||||||||| |||||||||||||||||||||||||||||||||
Db       3378 AGGACCGCGGCAGCCTTGGATGCACACGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 3437

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3438 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 3497

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       3498 CACCGGCTCGCGGTGGCGGCGACGTCGAGCGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 3557

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||| ||||||||| ||||||| ||||||||||||||||||||||||||
Db       3558 GCGCAGGGACAGACGCCGCCCGGTGTGGTGCGCGGTATCGCCGATTCCTCACGCGGCAAG 3617

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db       3618 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAT 3677

Qy       1681 GATGTATGGTCCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3678 GATGTATGGCCCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 3737

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3738 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 3797

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||
Db       3798 CTCGACCAGACAGCCTTTACCCAGCCGGCGCTGTTCACCTTCGAGTATGCGCTCGCCGCG 3857
```

```
Qy   1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
Db   3858 CTGTGGCGGTCGTGGGGCGTAGAGCCGGAGTTGGTCGCTGGCCATAGCATCGGTGAGCTG 3917

Qy   1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
Db   3918 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 3977

Qy   1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
Db   3978 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGCGGCGCCG 4037

Qy   2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
Db   4038 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 4097

Qy   2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
Db   4098 AACGGTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 4157

Qy   2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
Db   4158 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 4217

Qy   2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
Db   4218 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 4277

Qy   2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
Db   4278 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTGGCACAGACGAGGTGAGC 4337

Qy   2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
Db   4338 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 4397

Qy   2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
Db   4398 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 4457

Qy   2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
Db   4458 GGCCTGGTGCCTGCCTGCCTGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 4517

Qy   2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
Db   4518 GGGCGTGACGAGCCAGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 4577

Qy   2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
Db   4578 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 4637

Qy   2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
Db   4638 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 4697

Qy   2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
Db   4698 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGAAGGGGGGCGCGGTGCGCGGCGGC 4757

Qy   2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
Db   4758 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 4817

Qy   2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
Db   4818 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTCGATCGC 4877

Qy   2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
Db   4878 CTGGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTTGGCGAGGTCGAGATCGCCGTC 4937

Qy   2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
Db   4938 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 4997

Qy   3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
Db   4998 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 5057

Qy   3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
Db   5058 GCCGTGGGCGAGGGCGTGAACGGCCTTGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 5117

Qy   3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
Db   5118 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 5177

Qy   3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
Db   5178 CTCTCGGCGACCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 5237

Qy   3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
Db   5238 GACGGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCACGCGGCGACCGGCGGG 5297

Qy   3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
Db   5298 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 5357

Qy   3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
Db   5358 GGCACGCCCGAGAAGCGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 5417

Qy   3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
Db   5418 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 5477

Qy   3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
Db   5478 GTGCTCAACTCGCTTTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 5537

Qy   3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
Db   5538 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 5597

Qy   3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
Db   5598 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 5657

Qy   3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
Db   5658 CCGGCGCGGGTCCGTGCGCTCTTCGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 5717

Qy   3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
Db   5718 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCTCGTGTCGCCGATGCGTTCCGGAGCATG 5777

Qy   3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
Db   5778 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 5837

Qy   3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
Db   5838 CGTATTCCGACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGATCTGCTCGACAGG 5897

Qy   3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
Db   5898 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 5957

Qy   3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
Db   5958 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 6017

Qy   4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
```

```
Db       6018 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 6077

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140

Db       6078 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGACCCAAAAC 6137

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200

Db       6138 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 6197

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGAAATCATTGCC 4260

Db       6198 GCAGGCGTGCAAAGCGACTTCGTCTCATCGGGCGCAGATCAAGACTGGAAATCATTGCC 6257

Qy       4261 CTATGA 4266

Db       6258 CTATGA 6263


RESULT 9
US-09-266-965-19
; Sequence 19, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D
;  APPLICANT: Mao, Y
;  APPLICANT: Varoglu, M
;  APPLICANT: He, M
;  APPLICANT: Sheldon, P
;  TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
;  FILE REFERENCE: 600.456US1
;  CURRENT APPLICATION NUMBER: US/09/266,965
;  CURRENT FILING DATE: 1999-03-12
;  EARLIER APPLICATION NUMBER: US 08/624,447
;  EARLIER FILING DATE: 1996-08-19
;  EARLIER APPLICATION NUMBER: PCT/US94/11279
;  EARLIER FILING DATE: 1994-10-06
;  EARLIER APPLICATION NUMBER: US 08/133,963
;  EARLIER FILING DATE: 1993-10-07
;  NUMBER OF SEQ ID NOS: 145
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19
;   LENGTH: 3978
;   TYPE: DNA
;   ORGANISM: Streptomyces lavendulae
US-09-266-965-19

  Query Match             19.1%;  Score 814.2;  DB 4;  Length 3978;
  Best Local Similarity   57.6%;  Pred. No. 3.3e-134;
  Matches 1549;  Conservative    0;  Mismatches 1108;  Indels   30;  Gaps    4;

Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77

Db         90 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 149

Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137

Db        150 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 209

Qy        138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194

Db        210 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 269

Qy        195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254

Db        270 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 329

Qy        255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314

Db        330 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 389

Qy        315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374

Db        390 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 449

Qy        375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434

Db        450 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 509

Qy        435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGC 494

Db        510 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 563

Qy        495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554

Db        564 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 623

Qy        555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGAATGCTC 614

Db        624 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 683

Qy        615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674

Db        684 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 743

Qy        675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734

Db        744 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 803

Qy        735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794

Db        804 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 863

Qy        795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854

Db        864 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 923

Qy        855 CAGCGGTCTGACCGTGCCGAACGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914

Db        924 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 983

Qy        915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974

Db        984 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 1043

Qy        975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034

Db       1044 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGGCCCGCGGCGA 1103

Qy       1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094

Db       1104 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 1163

Qy       1095 GGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154

Db       1164 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 1223

Qy       1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214

Db       1224 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 1283

Qy       1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274

Db       1284 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 1343

Qy       1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334

Db       1344 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 1403

Qy       1335 ACCGCCG------------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376

Db       1404 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 1463

Qy       1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436

Db       1464 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 1523
```

```
Qy       1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496
Db       1524 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 1583

Qy       1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556
Db       1584 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 1643

Qy       1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616
Db       1644 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 1703

Qy       1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCT 1676
Db       1704 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCCGGCATGGGCCGCGAACT 1763

Qy       1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db       1764 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 1823

Qy       1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db       1824 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 1883

Qy       1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db       1884 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 1943

Qy       1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db       1944 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2003

Qy       1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
Db       2004 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 2063

Qy       1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db       2064 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 2123

Qy       2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db       2124 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 2183

Qy       2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
Db       2184 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 2243

Qy       2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216
Db       2244 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 2303

Qy       2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276
Db       2304 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 2363

Qy       2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336
Db       2364 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 2423

Qy       2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396
Db       2424 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 2483

Qy       2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456
Db       2484 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 2543

Qy       2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513
Db       2544 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 2603

Qy       2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGCTCTGGGCCGT 2573
Db       2604 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 2663

Qy       2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGCGGCGGGTGCCGCTGCC 2633
Db       2664 CGGCCTCTCGCCGGCGTGGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 2723

Qy       2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680
Db       2724 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 2770
```

RESULT 10
US-09-266-965-74
; Sequence 74, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09/266,965
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 74
;   LENGTH: 12249
;   TYPE: DNA
;   ORGANISM: Streptomyces lavendulae
US-09-266-965-74

```
  Query Match             19.1%;  Score 814.2;  DB 4;  Length 12249;
  Best Local Similarity   57.6%;  Pred. No. 3.5e-134;
  Matches 1549;  Conservative    0;  Mismatches 1108;  Indels   30;  Gaps    4;

Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77
Db       1077 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 1136

Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137
Db       1137 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 1196

Qy        138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194
Db       1197 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 1256

Qy        195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254
Db       1257 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 1316

Qy        255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314
Db       1317 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 1376

Qy        315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374
Db       1377 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 1436

Qy        375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434
Db       1437 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 1496
```

```
Qy   435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGC 494
Db  1497 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 1550

Qy   495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554
Db  1551 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 1610

Qy   555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614
Db  1611 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 1670

Qy   615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674
Db  1671 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 1730

Qy   675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734
Db  1731 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 1790

Qy   735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794
Db  1791 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 1850

Qy   795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854
Db  1851 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 1910

Qy   855 CAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914
Db  1911 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 1970

Qy   915 GGACGCAGGCTGCGCCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974
Db  1971 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 2030

Qy   975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034
Db  2031 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGCCCGCGGCGA 2090

Qy  1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094
Db  2091 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 2150

Qy  1095 GGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154
Db  2151 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 2210

Qy  1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214
Db  2211 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 2270

Qy  1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274
Db  2271 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 2330

Qy  1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334
Db  2331 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 2390

Qy  1335 ACCGCCG-----------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376
Db  2391 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 2450

Qy  1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436
Db  2451 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 2510

Qy  1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496
Db  2511 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 2570

Qy  1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556
Db  2571 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 2630

Qy  1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616
Db  2631 CCTGGCCAGCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 2690

Qy  1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCT 1676
Db  2691 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCCGGCATGGGCCGCGAACT 2750

Qy  1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db  2751 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 2810

Qy  1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db  2811 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 2870

Qy  1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db  2871 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 2930

Qy  1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db  2931 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2990

Qy  1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
Db  2991 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 3050

Qy  1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db  3051 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 3110

Qy  2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db  3111 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 3170

Qy  2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
Db  3171 CGTCAACGGACCGGAAGCGGTGGTGCTGTCGGCGACGAGGCCGCCGTCACGGACCTGGC 3230

Qy  2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216
Db  3231 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 3290

Qy  2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276
Db  3291 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 3350

Qy  2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336
Db  3351 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 3410

Qy  2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396
Db  3411 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 3470

Qy  2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456
Db  3471 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 3530

Qy  2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513
Db  3531 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 3590

Qy  2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGT 2573
Db  3591 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 3650

Qy  2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCC 2633
```

```
Db       3651 CGGCCTCTCGCCGGCGTGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 3710

Qy       2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680

Db       3711 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 3757


RESULT 11
US-09-266-965-96
; Sequence 96, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D
;  APPLICANT: Mao, Y
;  APPLICANT: Varoglu, M
;  APPLICANT: He, M
;  APPLICANT: Sheldon, P
;  TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
;  FILE REFERENCE: 600.456US1
;  CURRENT APPLICATION NUMBER: US/09/266,965
;  CURRENT FILING DATE: 1999-03-12
;  EARLIER APPLICATION NUMBER: US 08/624,447
;  EARLIER FILING DATE: 1996-08-19
;  EARLIER APPLICATION NUMBER: PCT/US94/11279
;  EARLIER FILING DATE: 1994-10-06
;  EARLIER APPLICATION NUMBER: US 08/133,963
;  EARLIER FILING DATE: 1993-10-07
;  NUMBER OF SEQ ID NOS: 145
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 96
;   LENGTH: 18331
;   TYPE: DNA
;   ORGANISM: Streptomyces lavendulae
US-09-266-965-96

  Query Match             19.1%;  Score 814.2;  DB 4;  Length 18331;
  Best Local Similarity   57.6%;  Pred. No. 3.6e-134;
  Matches 1549;  Conservative    0;  Mismatches 1108;  Indels   30;  Gaps    4;

Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77

Db       1077 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 1136

Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137

Db       1137 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 1196

Qy        138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194

Db       1197 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 1256

Qy        195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254

Db       1257 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 1316

Qy        255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314

Db       1317 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 1376

Qy        315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374

Db       1377 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 1436

Qy        375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434

Db       1437 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 1496

Qy        435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGC 494

Db       1497 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 1550

Qy        495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554

Db       1551 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 1610

Qy        555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614

Db       1611 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 1670

Qy        615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674

Db       1671 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 1730

Qy        675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734

Db       1731 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 1790

Qy        735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794

Db       1791 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 1850

Qy        795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854

Db       1851 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 1910

Qy        855 CAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914

Db       1911 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 1970

Qy        915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974

Db       1971 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 2030

Qy        975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034

Db       2031 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGCCCGCGGCGA 2090

Qy       1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094

Db       2091 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 2150

Qy       1095 GGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154

Db       2151 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 2210

Qy       1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214

Db       2211 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 2270

Qy       1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274

Db       2271 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 2330

Qy       1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334

Db       2331 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 2390

Qy       1335 ACCGCCG------------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376

Db       2391 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 2450

Qy       1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436

Db       2451 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 2510

Qy       1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496

Db       2511 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 2570

Qy       1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556

Db       2571 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 2630

Qy       1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616

Db       2631 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 2690
```

```
Qy       1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCT 1676
Db       2691 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCCGGCATGGGCCGCGAACT 2750

Qy       1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db       2751 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 2810

Qy       1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db       2811 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 2870

Qy       1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db       2871 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 2930

Qy       1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db       2931 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2990

Qy       1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
Db       2991 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 3050

Qy       1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db       3051 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 3110

Qy       2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db       3111 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 3170

Qy       2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
Db       3171 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 3230

Qy       2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216
Db       3231 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 3290

Qy       2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276
Db       3291 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 3350

Qy       2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336
Db       3351 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 3410

Qy       2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396
Db       3411 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 3470

Qy       2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456
Db       3471 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 3530

Qy       2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513
Db       3531 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 3590

Qy       2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGT 2573
Db       3591 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 3650

Qy       2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGCGGCGGGTGCCGCTGCC 2633
Db       3651 CGGCCTCTCGCCGGCGTGGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 3710

Qy       2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680
Db       3711 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 3757
```

RESULT 12
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144,085
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-144-085-3

```
  Query Match             18.8%;  Score 803.2;  DB 3;  Length 33529;
  Best Local Similarity   57.6%;  Pred. No. 3.1e-132;
  Matches 1591;  Conservative    0;  Mismatches 1128;  Indels   41;  Gaps    7;

Qy         37 CCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGG 96
Db      25316 CCGATAGCGATCGTGGGCGTGGGGCTGCGCATGCCTGGCGGAGCGAGCGACCTGGAGAGC 25375

Qy         97 TTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGG 156
Db      25376 TTCTGGCAGGTGCTGGTCGAAGAGCGGGATACGCTGCGGCCGATCCCGGCCCAACGATTC 25435

Qy        157 GATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGC 216
Db      25436 GACGTCGAGGCGCTGTACGATCCTGACCCCGACGCGAAGGGCAAGACGTACGTGCGGAAC 25495

Qy        217 GCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGC 276
Db      25496 GCGTCGCTGCTCGACGACGTGGCGTCGTTCGACCCTGGGTTCTTCGGGATAAGTCCGCGG 25555

Qy        277 GAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTG 336
Db      25556 GAGGCGGAGCCGATGGATCCGCAGCACCGGCTGCTGCTGGAGACGGCGTGGAGCGCCCTG 25615

Qy        337 GAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGG 396
Db      25616 GAGGACGCGGGGGTGCGTCCAGAGCACCTGAAGGGCTCGGACACGGGAGTGTTCGTGGGC 25675

Qy        397 ATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGAC 456
Db      25676 GTGGCGCCGAGCGAATACGCGAGCTACCGAGGAAA---------GAGCGCGAACGAAGAT 25726

Qy        457 GCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTC 516
Db      25727 GCGTATGCGCTGACAGGGACGGCGCTGAGCTTTGCTGCGGGACGGGTGGCCTACCACCTC 25786

Qy        517 GGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTT 576
Db      25787 GGGCTGCAAGGCCCTGCGGTGTCGACCGACACGGCCTGCAGCTCGTCGCTGGTAGCGGTG 25846

Qy        577 CATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTA 636
Db      25847 CACCTGGCGTGCGACGCGCTGCGCCGGGGCGATTGCGAGGTGGCGCTGGCGGCAGGTGTG 25906

Qy        637 TCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGG 696
Db      25907 CAGGTGCTGGCGAACCCGGCGGGGTTTGTGCTGCTGTCGCGCACGCGCGTTGTCGCCG 25966
```

```
Qy        697 GACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGC 756
Db      25967 GACGGGCGGTGCAAGGCGTTCTCGCAGGCGGCGGACGGTTATGGCCGTGGCGAGGGAGTC 26026

Qy        757 GCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCG 816
Db      26027 GGGGTGCTGGTGCTGATGCGGCTGTCCGAGGCGCAGCAGCAGGGGAAGCGGGTGCTGGGT 26086

Qy        817 GTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAAC 876
Db      26087 GTGGTGCGCGGCACGGCGGTCAATCAGGACGGCGCGAGCAGCGGGATCACGGCGCCGAAC 26146

Qy        877 GGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCT 936
Db      26147 GGCACGGCGCAGCAGAAGGTGGTGCGCGCGGCGCTGCGGAACGCGGGGCTGGAGCCGGCG 26206

Qy        937 TCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGACCCCATCGAAATC 996
Db      26207 AGCATCGATGTGGTGGAGTGCCACGGTACGGGCACGTCGCTGGGCGACCCGATCGAGGTG 26266

Qy        997 CAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGG 1056
Db      26267 CAGGCGCTCGGCGCGGTGTACGGGCAAGGTCGGGATATGGCTCGTCCGCTGCAGCTGGGC 26326

Qy       1057 TCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAG 1116
Db      26327 GCGGTCAAGAGCAATATCGGTCATCTCGAGTCCGCCGCGGGCATCGCAGGGGTGTGCAAG 26386

Qy       1117 GTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAAC 1176
Db      26387 ATCCTGGCGGCGTTCCGTTACGAGTCGCTGCCGGCGACGCTGCACAGCTCGCCGCGCAAT 26446

Qy       1177 CCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCGGACACCGTGGCC- 1235
Db      26447 CCCCGCATCCCGTGGGAGAACCTGCCGGTGCAGGTGGTCGATCGCCTGACCCCCTGGCCT 26506

Qy       1236 -----GGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGAGCGGGACC 1290
Db      26507 CGGCGCGCAGAGGGCCCCCCGCGCCGTGCCGGCGTGTCGTCGTTCGGGATCAGCGGGACG 26566

Qy       1291 AACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCC--------GG 1342
Db      26567 AACGCGCATGTCATCCTCGAAGAAGCGCCGGCCGAGGCGCGGCGCGAGCCTGTCGAGGCG 26626

Qy       1343 CGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGGATG 1402
Db      26627 AGGCGGCGCCTGCGCTATTGCCGCTGGTGCTGTCGGGTCGAGACGAGGCCGCGGTGAATG 26686

Qy       1403 CACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCGATG 1462
Db      26687 CGCAGGCGGGGCGGTGGGCGAAGTGGCTGGAAGAGCACGGGGAGGTGGGGTGGTCGGACG 26746

Qy       1463 TGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGGCGA 1522
Db      26747 TGGTGCGCACGGCGGCGCTGCACCGGACGCACTTCGAGTCGCGGGCGTCGGTGCTTGCGG 26806

Qy       1523 CGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGCCCG 1582
Db      26807 CGAGCGCTGCGGGAGCTGTGGAGGGTCTTCGCGCGCTGTCGTCGGGGCGGCCGGATGCGG 26866

Qy       1583 GTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAGCTCGCCTTTCTCTTCACCGGAC 1642
Db      26867 CGGTGGTGAGCGGGACGGCGAA---GCGAGGCGGGAAGCTTGCGGTGCTGTTCACGGGGC 26923

Qy       1643 AGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGTTCCGCG 1702
Db      26924 AGGGCAGCCAGCGGCTCGGGATGGGGAAGAGGCTTTACGAAGTGTACCCCGTGTTCCGTG 26983

Qy       1703 AGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCCGCGAGG 1762
Db      26984 CGGCGTTCGACGAGGTGTGCGAGGCGCTGGACGCGCATCTCGACCGTGGGTTGAGAGAGG 27043

Qy       1763 TGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCTTCACCC 1822
Db      27044 TGGTGTTCGCGGCCGCGGGCAGCGAGGAAGGAGCGCAGCTGGAGCGGACGGAGTACACGC 27103

Qy       1823 AGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAG 1882
Db      27104 AGCCCGGCTGTTTGCGCTGGAAGTGGCGCTGTACCGTCAGTGGGAGTCGTGGGGGCTGA 27163

Qy       1883 AGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCG 1942
Db      27164 AGCCCGCTGCGCTKCTGGGGCACTCGATAGGAGAGCTGAGCGCTGCGCACGTGGCGGGTG 27223

Qy       1943 TGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGC 2002
Db      27224 TGCTGAGCCTTGCGGACGCAGCGAAGCTAGTGTGCGCCCGCGGTCGGCTGATGCAGGGGT 27283

Qy       2003 TGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGG 2062
Db      27284 GCGAGGCCGGGGGAGCGATGGTGTCGGTGGAGGCCTCGGAGCCGGAGGTGCAGCGGGCGC 27343

Qy       2063 TGGCGCCGCACGCAGCGTCGG------TGTCGATCGCCGCGGTCAACGCTCCGGACCAGG 2116
Db      27344 TGTCGGAGGTCGGGGCGCAGGGGCGACTGAGCATCGCCGGGCTGAACGCGCCGATGCAGA 27403

Qy       2117 TGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCG 2176
Db      27404 CGGTGCTGAGCGGGGACGAAGCGGCGGTGCTCGCGGTGGCGCGACGGCTGGAGGCGCAGG 27463

Qy       2177 GGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGA 2236
Db      27464 GCCGGCGCACGCGGCGTCTGCGTGTGTCGCACGCGTTCCACAGCGCGCACATGGACGGGA 27523

Qy       2237 TGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCC 2296
Db      27524 TGCTGGAGGAGTTCGGGAAGGTGGCGCGGGAGTGCACGTACGCGCGGCCGCAGCTGGCGG 27583

Qy       2297 TGGTCAGCAATCTGA---GCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATT 2353
Db      27584 TGGTGAGCGGCGTGACGGGCGAGCTCGGTGGCGAAGAAGCGCTGATGTCGGCCGAGTACT 27643

Qy       2354 GGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGG 2413
Db      27644 GGGTGAGGCAGGTGCGCGAGGCGGTGCGCTTCCTGGACGGGATGCGCACGCTTGCGGCGG 27703

Qy       2414 CCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCTGGTGCCTG 2473
Db      27704 CGGGGGTGAGCACATACGTCGAGTGTGGGCCGGATGGCGTGCTGTGCGCGCTGGGGGCGG 27763

Qy       2474 CCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATC------GTCGCGCGCTGGGCGTG 2527
Db      27764 GGTGCCTGCCGGAGGGAGCCGAGGCGACGTTTGTGGCGAGCCTGCGGCGAGAGCAGGAGG 27823

Qy       2528 ACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCT 2587
Db      27824 AAGAGCGCGCGCTGGCGACGGCGGTGGCGACAGTGCACGTGCAGGGGCACGAGGTGGACT 27883

Qy       2588 CCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGC 2647
Db      27884 GGGCCCAGGTGCTGTCGGGCCGTGGCGGCCGGCCCGTGGAGCTGCCGACGTACGCGTTCC 27943

Qy       2648 AGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGACCGCCGTG 2707
Db      27944 AGCGGCAGCGCTACTGGCTGGAAGCGCCGAAGGCGCGTACCGACGTGGCTCGGCGGGCT 28003

Qy       2708 CTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCGGC 2767
Db      28004 TGAGGGAGTCGGGGCATCCGCTGCTCGGAGCGGCAACGAAGCTGGCCGACGGCGACGGCC 28063
```

RESULT 13
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991

```
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             17.7%;  Score 755.2;  DB 2;  Length 43280;
  Best Local Similarity   56.0%;  Pred. No. 7.8e-124;
  Matches 1603;  Conservative    0;  Mismatches 1208;  Indels   53;  Gaps    7;

Qy         24 CGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGAT 83
Db      36350 CGCGGCGCGGGAGCCGATCGCCGTCGTCGGCATGGCCTGCCGCTACCCGGGCGGTGTCAC 36409

Qy         84 CGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCC 143
Db      36410 CGGCCCCGAGGAGCTGTGGGAGCTGGTGGCCGGAGGCCGGGACGCGATCGGGCCGTTCCC 36469

Qy        144 CGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAA 200
Db      36470 CGTGGACCGGGGCTGGGACGTGGCGTCGGTGTACGACCCGGATCCCGAGTCGAAGGGCAC 36529

Qy        201 GACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTT 260
Db      36530 CACGTACTGCCGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTT 36589

Qy        261 CGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGT 320
Db      36590 CGGCATCTCGCCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGT 36649

Qy        321 GTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAAC 380
Db      36650 GTCCTGGGAGGCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGG 36709

Qy        381 GGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGC 440
Db      36710 TGGTGTCTACGTGGGCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGA 36769

Qy        441 GTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCG 500
Db      36770 GGGCTCGGAGGGCTATCTGCTGACCGGCAGCGCCGACGCG------GTGATGTCCGGCCG 36823

Qy        501 AATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTC 560
Db      36824 CATCTCCTACGCGCTCGGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTC 36883

Qy        561 CTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGC 620
Db      36884 CTCGCTGGTGGCGCTGCATCTGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGC 36943

Qy        621 CCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGAC 680
Db      36944 GCTGGCGGGCGGGGTGGCGGTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCA 37003

Qy        681 CCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGG 740
Db      37004 GAAGGGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGG 37063

Qy        741 ACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGG 800
Db      37064 CTGGGCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGG 37123

Qy        801 CGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGG 860
Db      37124 GCACACGGTCCTCGGCCTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGG 37183

Qy        861 TCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGC 920
Db      37184 GCTGACCGCGCCCAACGGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGC 37243

Qy        921 AGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGG 980
Db      37244 CGGGCTGTCCCCGGAGGACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGG 37303

Qy        981 TGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCAC 1040
Db      37304 CGACCCCATCGAGGCCGGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCA 37363

Qy       1041 GCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGAT 1100
Db      37364 CCCGCTGTGGCTCGGCTCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGT 37423

Qy       1101 CACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCA 1160
Db      37424 CGGCGGCGTCATCAAGATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCA 37483

Qy       1161 CGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGC 1220
Db      37484 CGCCGACGAGCCGACCCCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTC 37543

Qy       1221 CCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCAT 1280
Db      37544 CGAGGTGCCGTGGCAGCGGACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGT 37603

Qy       1281 GAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCC 1340
Db      37604 CGGCGGCACCAATGCCCATGTCGTCCTCGAAGAGGCACCCGCCCCGCCCGCGCCGGAACC 37663

Qy       1341 GGCGCCGGAGCGACC--------------------------------GGCAGAGCTGCT 1367
```

```
Db      37664 GGCCGGGGAGGCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGGCCCCTGGCCTG 37723

Qy       1368 GGTGCTGTCGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCA 1427
Db      37724 GGTGGTCTCCGGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCA 37783

Qy       1428 TCTGGAGACCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCG 1487
Db      37784 CCTCTCCCGCACCCCCGGGGCCCGCCCGCCGTGACATCGCCTTCTCCCTCGCCGCCACGCG 37843

Qy       1488 CAGCGCGATGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGC 1547
Db      37844 CGCAGCCTTTGACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGC 37903

Qy       1548 CCTGGACGCTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTC 1607
Db      37904 CCTGGACGCGTTGGCCGAAGGACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGA--- 37960

Qy       1608 CTCACGCGGCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGG 1667
Db      37961 CCGGGACGGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCCGGGATGGC 38020

Qy       1668 CCGTGGGCTGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCT 1727
Db      38021 CCACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTGACGGACCG 38080

Qy       1728 GTTCAACCAGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGT 1787
Db      38081 TCTCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGCTGGACGCCCGACCCGGCTCGCC 38140

Qy       1788 CGACGCCGCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATA 1847
Db      38141 CGAAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCGGCGCTCTTCGCCGTCGAGGT 38200

Qy       1848 TGCGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAG 1907
Db      38201 GGCGCTCCACCGGCTGCTGGAGCACTGGGGATGCGCCCCGACCTGCTGCTGGGGCACTC 38260

Qy       1908 CATCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTT 1967
Db      38261 GGTGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTGCTCGATCTCGACGACGCCTGCGC 38320

Qy       1968 CCTGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTC 2027
Db      38321 GCTGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTGCCGCCCGGCGGCGCGATGGTCTC 38380

Qy       2028 GATCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTC 2087
Db      38381 CGTGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTGGCCGGCCGCGAGGACGCCGTCTG 38440

Qy       2088 GATCGCCGCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCA 2147
Db      38441 CGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGC 38500

Qy       2148 TGCGATCGCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCA 2207
Db      38501 CGAGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCA 38560

Qy       2208 TGCGTTCCACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTC 2267
Db      38561 CGCCTTCCACTCACCCCTGATGGACGGCATGCTCGCCGGATTCCGGGAGGTCGCCGCCGG 38620

Qy       2268 GGTGAGCTACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAG---GCTTG 2324
Db      38621 CCTGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCCACGGTCACGGGGCGGCCCGCCCG 38680

Qy       2325 CACAGACGAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTT 2384
Db      38681 CCCCGGTGAACTCACCGGCCCCGACTACTGGGTGGCCCAGGTCCGTGAGCCCGTGCGCTT 38740

Qy       2385 CGCGGATGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCC 2444
Db      38741 CGCGGACGCGGTCCGCACGGCACACCGCCTCGGAGCCCGCACCTTCCTGGAGACCGGCCC 38800

Qy       2445 GAAATCGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCT 2504
Db      38801 GGACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTGGAGGACGACACCGTGGCCCTGCT 38860

Qy       2505 CGCATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCT 2564
Db      38861 ---GCCGGCGATCCACAAGCCCGGCACCGCGCCGCACGGTCCGGCGGCTCCCGGCGCGCT 38917

Qy       2565 CTGGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGCGGCGGGT 2624
Db      38918 GCGGGCGGCCGCCGCCGCGTACGGCCGGGGCGCCCGGGTGGACT--GGGCCGGGATGCAC 38975

Qy       2625 GCCGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGA 2684
Db      38976 GCCGACGGCCCCGAGGGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTTCCGGCAC 39035

Qy       2685 CGCGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGG 2744
Db      39036 CGCCGCTACTGGCTCGCCCCGGGCCGCGCGGCGGACACCGACGACTGGATGTACCGGATC 39095

Qy       2745 CGCGGTGCGCGGCGGCGACCGGCGCAGCGCTCGGCTCGACCATCCCGCCGCCCGAGAGCGG 2804
Db      39096 GGCTGGGACCGGCTGCCGGCTGTGACCGGCGGGGCCCGGACCGCCGGCCGCTGGCTGGTG 39155

Qy       2805 ACGCCGGGAGAAGGTCGAGGCCGCCGGCGACCGTCCGTTCCGGC 2848
Db      39156 ATCCACCCCGACAGCCCGCGCTGCCGGGAGCTGTCCGGCCACGC 39199
```

RESULT 14
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match   17.4%; Score 742; DB 3; Length 15872;
  Best Local Similarity   55.7%; Pred. No. 1.5e-121;
  Matches 1594; Conservative   0; Mismatches 1215; Indels   53; Gaps   7;

```
Qy         26 CAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCG 85
Db       8490 CCGACGACGATCCGATCGCGATCGTGGGCATGGCATGCCGCTATCCGGGTGGTGTGTCGT 8549

Qy         86 ATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCG 145
Db       8550 CGCCGGAGGACCTGTGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCG 8609

Qy        146 CCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGA 202
Db       8610 TCAACCGCGGCTGGGACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCA 8669

Qy        203 CGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCG 262
Db       8670 CGTACTGCCGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCG 8729

Qy        263 GCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGT 322
```

```
Db     8730 GCATCTCGCCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGT 8789

Qy      323 GCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGG 382
Db     8790 CCTGGGAGGCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTG 8849

Qy      383 GAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGT 442
Db     8850 GTGTCTACGTGGGCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGG 8909

Qy      443 CCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAA 502
Db     8910 GCTCGGAG------GGCTATCTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCA 8963

Qy      503 TCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCT 562
Db     8964 TCTCCTACGCGCTCGGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCT 9023

Qy      563 CGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCC 622
Db     9024 CGCTGGTGGCGCTGCATCTGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGC 9083

Qy      623 TGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCC 682
Db     9084 TGGCGGGCGGGGTGGCGGTGATGGCCGATCGGCGGCGTTCGTGGAGTTCTCCCGGCAGA 9143

Qy      683 GGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGAC 742
Db     9144 AGGGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCT 9203

Qy      743 GAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCG 802
Db     9204 GGGCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGC 9263

Qy      803 ATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTC 862
Db     9264 ACACGGTCCTCGGCCTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGC 9323

Qy      863 TGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAG 922
Db     9324 TGACCGCGCCCAACGGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCGACGCCG 9383

Qy      923 GCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTG 982
Db     9384 GGCTGTCCCCGGAGGACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCG 9443

Qy      983 ACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGC 1042
Db     9444 ACCCCATCGAGGCCGGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACC 9503

Qy     1043 CGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCA 1102
Db     9504 CGCTGTGGCTCGGCTCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCG 9563

Qy     1103 CTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACG 1162
Db     9564 GCGGCGTCATCAAGATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACG 9623

Qy     1163 CGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCC 1222
Db     9624 CCGACGAGCCGACCCCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCG 9683

Qy     1223 GGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGA 1282
Db     9684 AGGTGCCGTGGCAGCGGACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCG 9743

Qy     1283 GCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCGG 1342
Db     9744 GCGGCACCAATGCCCATGTCGTCCTCGAAGAGGCACCCGCCCCGCCCGCGCCGGAACCGG 9803

Qy     1343 CGCCGGAGCGACC---------------------------------GGCAGAGCTGCTGG 1369
Db     9804 CCGGGGAGGCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGCCCCTGGCCTGGG 9863

Qy     1370 TGCTGTCGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATC 1429
Db     9864 TGGTCTCCGGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCACC 9923

Qy     1430 TGGAGACCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCA 1489
Db     9924 TCTCCCGCACCCCCGGGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCGCCACGCGCG 9983

Qy     1490 GCGCGATGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCC 1549
Db     9984 CAGCCTTTGACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGCCC 10043

Qy     1550 TGGACGCTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCT 1609
Db    10044 TGGACGCGTTGGCCGAAGGACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGA---CC 10100

Qy     1610 CACGCGGCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCC 1669
Db    10101 GGGACGGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCCGGGATGGCCC 10160

Qy     1670 GTGGGCTGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGT 1729
Db    10161 ACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTGACGGACCGTC 10220

Qy     1730 TCAACCAGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCG 1789
Db    10221 TCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGGACGCCCGACCCGGCTCGCCCG 10280

Qy     1790 ACGCCGCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATG 1849
Db    10281 AAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCGGCGCTCTTCGCCGTCGAGGTGG 10340

Qy     1850 CGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCA 1909
Db    10341 CGCTCCACCGGCTGCTGGAGCACTGGGGGATGCGCCCCGACCTGCTGCTGGGGCACTCGG 10400

Qy     1910 TCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCC 1969
Db    10401 TGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTGCTCGATCTCGACGACGCCTGCGCGC 10460

Qy     1970 TGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGA 2029
Db    10461 TGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTGCCGCCCGGCGGCGCGATGGTCTCCG 10520

Qy     2030 TCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGA 2089
Db    10521 TGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTGGCCGGCCGCGAGGACGCCGTCTGCG 10580

Qy     2090 TCGCCGCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATG 2149
Db    10581 TCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCG 10640

Qy     2150 CGATCGCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATG 2209
Db    10641 AGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCACG 10700

Qy     2210 CGTTCCACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGG 2269
Db    10701 CCTTCCACTCACCCCTGATGGACGGCATGCTCGCCGGATTCCGGGAGGTCGCCGCCGGCC 10760

Qy     2270 TGAGCTACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAG---GCTTGCA 2326
Db    10761 TGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCCACGGTCACGGGGCGGCCCGCCCGCC 10820

Qy     2327 CAGACGAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCG 2386
Db    10821 CCGGTGAACTCACCGGCCCCGACTACTGGGTGGCCCAGGTCCGTGAGCCCGTGCGCTTCG 10880

Qy     2387 CGGATGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGA 2446
Db    10881 CGGACGCGGTCCGCACGGCACACCGCCTCGGAGCCCGCACCTTCCTGGAGACCGGCCCGG 10940
```

```
Qy       2447 AATCGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCG 2506
Db      10941 ACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTGGAGGACGACACCGTGGCCCTGCTGC 11000

Qy       2507 CATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCT 2566
Db      11001 ---CGGCGATCCACAAGCCCGGCACCGCGCCGCACGGTCCGGCGGCTCCCGGCGCGCTGC 11057

Qy       2567 GGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGC 2626
Db      11058 GGGCGGCCGCCGCCGCGTACGGCCGGGGCGCCCGGGTGGACT--GGGCCGGGATGCACGC 11115

Qy       2627 CGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACG 2686
Db      11116 CGACGGCCCCGAGGGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTTCCGGCACCG 11175

Qy       2687 CGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCG 2746
Db      11176 CCGCTACTGGCTCGCCCCGGGCCGCGCGGCGGACACCGACGACTGGATGTACCGGATCGG 11235

Qy       2747 CGGTGCGCGGCGGCGACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGAC 2806
Db      11236 CTGGGACCGGCTGCCGGCTGTGACCGGCGGGGCCCGGACCGCCGGCCGCTGGCTGGTGAT 11295

Qy       2807 GCCGGGAGAAGGTCGAGGCCGCCGGCGACCGTCCGTTCCGGC 2848
Db      11296 CCACCCCGACAGCCCGCGCTGCCGGGAGCTGTCCGGCCACGC 11337
```

RESULT 15
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

```
  Query Match            17.4%;  Score 740.8;  DB 3;  Length 4411529;
  Best Local Similarity   56.7%;  Pred. No. 3.3e-121;
  Matches 1571;  Conservative    0;  Mismatches 1127;  Indels   74;  Gaps    8;

Qy         19 GAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGC 78
Db    3297714 GAGCAACGCGCGACAGAACCGGTGGCGGTGGTGGGAATCGGCTGCCGATTTCCGGGCGGC 3297655

Qy         79 GTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGA 138
Db    3297654 GTGGATGGCCCGGACGGACTCTGGGATGTCGTTTCGGCGGGGCGCGACGTGGTGTCGGAG 3297595

Qy        139 GTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCG 195
Db    3297594 TTTCCGACCGATCGGGGCTGGGATGTTGAGGGGCTATACGATCCGGATCCCGACGCTGAG 3297535

Qy        196 GGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCC 255
Db    3297534 GGCAAAACCTACACGCGATGGGGGCATTTCTAGACGACGCCACCGGCTTTGACGCCGGG 3297475

Qy        256 TTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTG 315
Db    3297474 TTCTTCGGCATCGCTCCTAGTGAAGTGTTGGCGATGGATCCTCAGCAGCGGTTGATGCTG 3297415

Qy        316 GAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACG 375
Db    3297414 GAGGTTTCGTGGAGGCGTTGGAGCACGCGGGGATTGACCCGTTGTCGTTGCGTGGTTCG 3297355

Qy        376 GAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCG 435
Db    3297354 GCGACCGGGGTGTACACCGGCATTTTCGCGGCGAGTTACG--------------GCAAT 3297310

Qy        436 ACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCG 495
Db    3297309 AGGGATACCGGGGGGCTGCAAGGGTACGGGTTGACCGGCACGTCAATCAGCGTGGCCTCG 3297250

Qy        496 GGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTAT 555
Db    3297249 GGACGTGTGTCGTACGTATTGGGGCTGCAGGGCCCGGCGGTGTCAGTGGATACCGCTTGT 3297190

Qy        556 TCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCC 615
Db    3297189 TCGTCGTCGTTGGTGGCGATCCATTGGGCGATGTCGTCATTGCGCTCCGGGGAGTGTGAT 3297130

Qy        616 ACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCG 675
Db    3297129 TTGGCGTTGGCCGGTGGGGTGACGGTGATGGGGCTGCCGTCGATTTTTGTGGGGTTTTCC 3297070

Qy        676 AAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGG 735
Db    3297069 CGGCAGCGGGGGTTGGCTGCCGATGGGCGATGTAAGGCGTTTGCGGCCGCGGCCGACGGC 3297010

Qy        736 TTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCG 795
Db    3297009 ACCGGGTGGGGCGAGGGCGCCGGGGTGGTTGTGCTGGAGCGGCTATCGGACGCCCGGCGG 3296950

Qy        796 GACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGC 855
Db    3296949 TTGGGGCATTCGGTGCTGGCGGTGGTGCGGGGGAGCGCGGTCAACCAGGATGGTGCGTCC 3296890

Qy        856 AGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCG 915
Db    3296889 AATGGGCTGACCGCGCCCAATGGGCTCGCCCAGCAGCGAGTGATCCAGGTGGCGTTGGCC 3296830

Qy        916 GACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACG 975
Db    3296829 AACGCCGGGTTGTCCGCCGCAGACGTAGACGTGGTGGAGGCACACGGGACCGCCACCACG 3296770

Qy        976 CTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTC 1035
Db    3296769 TTGGGTGATCCCATCGAAGCGCAGGCATTGCTGTCCACCTACGGGCAGGGAGGTCCAGCG 3296710

Qy       1036 GCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCG 1095
Db    3296709 GAGCAGCCGCTGTGGGTGGGCTCAATCAAGTCGAACATGGGCCACACCCAGGCCGCCGCG 3296650

Qy       1096 GGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCAC 1155
Db    3296649 GGGGTCGCCGGGGTGATCAAGATGGTGCAGGCGATGCGTCATGGGGTGATGCCGGCGACG 3296590

Qy       1156 CTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACG 1215
Db    3296589 TTGCATGTGGATGAGCCTTCCCCGCGGGTGGATTGGACCAGCGGCGCGGTGTCGGTGTTG 3296530

Qy       1216 CGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTC 1275
Db    3296529 ACCGAGGCCAGGGAGTGGTCGGTTGACGGGCGTCCGCGCCGGGCGGCGGTGTCCTCCTTT 3296470

Qy       1276 GGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGAC------- 1328
Db    3296469 GGGATCAGCGGCACCAACGCGCATCTGATCTTGGAAGAGGCCCCGTACCAGCCCCCGCA 3296410

Qy       1329 ------------------GTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGG 1369
```

```
Db   3296409 GAGGCGCCGGTTGAGGCGAGCGAAAGCACCGGGGGCCGCGGCCGTCGATGGTGCCGTGGG 3296350

Qy      1370 TGCTGTCGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATC 1429
Db   3296349 TGATTTCGGCGAGATCGGCTGAGGCGTTGACGGCGCAGGCGGGTCGACTTATGGCCCACG 3296290

Qy      1430 TGGAGACCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCA 1489
Db   3296289 TGCAGGCCAACCCAGGGCTGGATCCGATCGATGTGGGGTGCTCGTTGGCCAGT---CGCT 3296233

Qy      1490 GCGCGATGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCC 1549
Db   3296232 CGGTGTTTGAGCACCGAGCGGTGGTGGTCGGCGCAAGCCGTGAGCAACTGATTGCCGGGC 3296173

Qy      1550 TGGACGCTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCT 1609
Db   3296172 TGGCTGGGCTCGCGGCGGGCGAGCCGGGTGCCGGCGTGGCGGTCGGTCAGCCAG---GGT 3296116

Qy      1610 CACGCGGCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCC 1669
Db   3296115 CGGTGGGCAAGACGGTGGTCGTGTTTCCTGGGCAGGGCGCGCAGCGCATCGGGATGGGCC 3296056

Qy      1670 GTGGGCTGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGT 1729
Db   3296055 GCGAGTTGTACGGCGAGTTGCCCGTGTTTGCGCAGGCATTCGATGCGGTGGCCGACGAGT 3295996

Qy      1730 TCAACCAGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCG 1789
Db   3295995 TGGACCGGCATCTGCGGTTGCCGCTGCGCGACGTTATTTGGG-----------GTGCCG 3295948

Qy      1790 ACGCCGCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATG 1849
Db   3295947 ATGCGGATTTGCTTGACAGCACCGAATTTGCTCAGCCCGCGTTGTTCGCGGTGGAGGTGG 3295888

Qy      1850 CGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCA 1909
Db   3295887 CATCGTTCGCGGTGTTGCGGGATTGGGGTGTGCTTCCGGACTTCGTCATGGGTCACTCCG 3295828

Qy      1910 TCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCC 1969
Db   3295827 TTGGAGAGCTGGCGGCGGCGCACGCGGCCGGTGTGTTGACGTTGGCGGACGCGGCGATGC 3295768

Qy      1970 TGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGA 2029
Db   3295767 TGGTGGTGGCGCGGGGCCGGTTGATGCAGGCGCTGCCGGCAGGCGGTGCGATGGTGGCGG 3295708

Qy      2030 TCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGA 2089
Db   3295707 TGGCTGCCAGTGAGGACGAGGT---------GGAGCCGCTGCTGGGTGAGGGTGTGGGGA 3295657

Qy      2090 TCGCCGCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATG 2149
Db   3295656 TCGCTGCGATCAACGCGCCCGAATCGGTGGTGATCTCCGGTGCGCAGGCCGCGGCAAATG 3295597

Qy      2150 CGATCGCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATG 2209
Db   3295596 CGATTGCGGATCGGTTCGCCGCGCAGGGTCGGCGGGTGCACCAGTTGGCGGTCTCGCATG 3295537

Qy      2210 CGTTCCACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGG 2269
Db   3295536 CGTTTCATTCGCCGTTGATGGAGCCGATGCTCGAGGAGTTCGCGCGTGTCGCGGCCCGGG 3295477

Qy      2270 TGAGCTACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAG 2329
Db   3295476 TGCAGGCACGCGAGCCCCAGCTTGGGCTGGTGTCGAACGTGACGGGCGAGTTGGCCGGCC 3295417

Qy      2330 ACGAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGG 2389
Db   3295416 CTGATTTCGGGTCGGCGCAGTACTGGGTGGACCACGTTCGTCGGCCGGTGCGCTTCGCGG 3295357

Qy      2390 ATGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAAT 2449
Db   3295356 ACAGTGCGCGTCATTTGCAGACCCTTGGGGCGACCCACTTCATCGAGGCCGGCCCGGGAA 3295297

Qy      2450 CGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCAT 2509
Db   3295296 GTGGTTTGACTGGCTCGATCGAGCAGTCCTTGGCCCCGGCTGAGGCGATGGTGGTGTCGA 3295237

Qy      2510 CGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGG 2569
Db   3295236 TGCTGGGCAAAGACCGGCCCGAGCTGGCCTCGGCGCTCGGTGCTGCCGGTCAGGTGTTCA 3295177

Qy      2570 CCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCC---CTCAGGGGGGCGGCGGGTGC 2626
Db   3295176 CCACCGGTGTGCCGGTGCAGTGGTCGGCGGTGTTCGCCGGCTCGGGTGGACGGCGGGTGC 3295117

Qy      2627 CGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACG 2686
Db   3295116 AGCTGCCCACGTATGCGTTTCAGCGACGGCGGTTTTGGGAGACGCCGGGCGCGGATGGGC 3295057

Qy      2687 CGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCG 2746
Db   3295056 CCGCCGATGCGGCCGGGTTGGGTCTGGGCGCGACCGAGCATGCCTTGTTGGTGCGGTGG 3294997

Qy      2747 CGGTGCGCGGCG 2758
Db   3294996 TCGAGCGGCCCG 3294985
```

Search completed: October 4, 2003, 03:39:32
Job time : 275.999 secs

```
OM nucleic - nucleic search, using sw model

Run on:         October 3, 2003, 11:29:26 ; Search time 1043.88 Seconds
                                            (without alignments)
                                            11031.749 Million cell updates/sec

Title:          US-10-014-717-1_COPY_7610_11875
Perfect score:  4266
Sequence:       1 gtggcggatcgtcccatcga..........gggaaatcattgccctatga 4266

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters:      5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_19Jun03:*
                 1:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                 2:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                 3:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                 4:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                 5:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                 6:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                 7:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                 8:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                 9:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                 10:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                 11:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                 12:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                 13:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                 14:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                 15:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                 16:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                 17:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                 18:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                 19:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
                 20:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                 21:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                 22:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                 23:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                 24:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
                 25:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4266	100.0	68750	21	AAZ55887	Sorangium cellulos
	2	4187.6	98.2	71989	21	AAA29349	Sorangium cellulos
	3	817.2	19.2	67459	25	ABQ77491	S. aurantiaca DNA
	4	814.2	19.1	3978	21	AAC55785	Type I polyketide
	5	814.2	19.1	12249	21	AAC55840	Complete Mitomycin
	6	814.2	19.1	18331	21	AAC55857	Complete nucleotid
	7	805	18.9	27541	22	AAD17185	Streptomyces nours
	8	805	18.9	125401	22	AAD17186	Streptomyces nours
	9	803.2	18.8	33529	23	AAS17367	DNA sequence of S.
	10	766.6	18.0	20394	22	AAF24892	Pimaricin biosynth
	11	755.2	17.7	43280	18	AAT80413	Tylactone synthase
	12	745.8	17.5	53789	19	AAV21187	Amycolatopsis medi
	13	742	17.4	15872	21	AAZ87283	S. venezuelae vep
c	14	740.8	17.4	4411529	22	AAI99682	Mycobacterium tube
	15	735.6	17.2	15872	18	AAT68715	Streptomyces venez
c	16	732	17.2	4403765	22	AAI99683	Mycobacterium tube
	17	726.2	17.0	65140	22	AAD17184	Streptomyces nours
	18	725	17.0	49377	19	AAV05287	The soraphen biosy
	19	701.2	16.4	16767	22	AAF88339	S. spinosa DNA fra
	20	701.2	16.4	50000	22	AAF88313	S. spinosa DNA fra
	21	701.2	16.4	80161	20	AAZ21501	DNA fragment of Sa
	22	679.2	15.9	14775	22	AAF88338	S. spinosa DNA fra
	23	679.2	15.9	50000	22	AAF88316	S. spinosa DNA fra
	24	677.2	15.9	75236	25	ABV75557	Saccharopolyspora
	25	675.2	15.8	44377	18	AAT78508	Platenolide syntha
	26	675.2	15.8	44377	18	AAT80414	Platenolide syntha
	27	671.8	15.7	28598	17	AAT06769	Sorangium cellulos
	28	671.8	15.7	28958	18	AAT89956	Sorangium cellulos
	29	671.8	15.7	28958	21	AAA75299	DNA sequence of So
	30	671	15.7	9513	22	AAF88337	S. spinosa DNA fra
	31	663.2	15.5	29879	14	AAQ46806	eryA region of S.
	32	662.2	15.5	13987	18	AAT80415	Hybrid srmG/tylG O
	33	649.4	15.2	47981	22	AAF30757	Micromonospora meg
	34	639	15.0	6459	22	AAF88336	S. spinosa DNA fra
	35	633.2	14.8	5088	22	AAF90038	Nucleotide sequenc
	36	633.2	14.8	34071	22	AAF90033	Nucleotide sequenc
c	37	633.2	14.8	42717	22	AAF90032	Nucleotide sequenc
	38	633	14.8	77536	21	AAA14651	Nucleotide sequenc
	39	628.8	14.7	4403765	22	AAI99683	Mycobacterium tube
	40	628.8	14.7	4411529	22	AAI99682	Mycobacterium tube
	41	627.6	14.7	13842	21	AAZ87297	S. venezuelae macr
	42	627.6	14.7	36778	21	AAZ87318	S. venezuelae pik
	43	627.6	14.7	37948	21	AAZ87285	S. venezuelae pik
	44	627.6	14.7	38506	21	AAA75633	Nucleotide sequenc
	45	627.6	14.7	38506	21	AAZ56001	Recombinant cosmid

ALIGNMENTS

```
RESULT 1
AAZ55887
ID   AAZ55887 standard; DNA; 68750 BP.
XX
AC   AAZ55887;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum 68.75 kb contig.
XX
KW   Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW   anticancer; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1826
FT                   /*tag=  a
FT                   /partial
FT                   /product= "Partial Orf 1 protein (AAY58580)"
FT                   /note= "No initiation codon given in the specification"
FT   CDS             complement (1900..3171)
FT                   /*tag=  b
FT                   /product= "Orf 2 protein (AAY58581)"
FT   CDS             3415..5556
FT                   /*tag=  c
FT                   /product= "Orf 3 protein (AAY58582)"
FT   CDS             complement (5612..5992)
FT                   /*tag=  d
FT                   /product= "Orf 4 protein (AAY58583)"
FT   CDS             6226..6675
```

```
FT                   /*tag= e
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                             (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                             (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                             (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                             (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                             (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                             (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                             (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR   AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR   AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR   AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy  -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             100.0%;  Score 4266;  DB 21;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy         421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy         481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy         541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy         601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy         661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy         721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy         781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy         841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy         901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy         961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy        1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy        1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy        1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy        1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy        1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy        1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy        1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy        1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy        1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy        1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy        1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy        1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy        1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy        1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy        1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy        1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy        1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy        2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy        2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy        2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy        2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy        2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy        2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy        2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy        2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy        2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189
```

```
Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875
```

```
RESULT 2
AAA29349
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..992
FT                   /*tag=  a
FT                   /label= ORF_A
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             989..1501
FT                   /*tag=  b
FT                   /label= ORF_B
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             1998..6263
```

```
FT                   /*tag=  c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031..3548
FT                   /*tag=  d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                          domain"
FT   misc_RNA        3621..4661
FT                   /*tag=  e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                          domain"
FT   misc_RNA        4917..5810
FT                   /*tag=  f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                          domain, potentially involved in formation of the
FT                          thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag=  g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                          domain"
FT   CDS             6260..10493
FT                   /*tag=  h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag=  i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                          module"
FT   misc_RNA        2031..3548
FT                   /*tag=  j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                          module"
FT   misc_RNA        6861..6887
FT                   /*tag=  k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag=  l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                          module"
FT   misc_RNA        7358..7366
FT                   /*tag=  m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                          NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag=  n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag=  o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag=  p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag=  q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag=  r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag=  s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag=  t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag=  u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag=  v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                          module"
FT   misc_RNA        10261..10306
FT                   /*tag=  w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                          module"
FT   CDS             10639..16137
FT                   /*tag=  x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag=  y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag=  z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag=  aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                          module 2"
FT   misc_RNA        14962..15756
FT                   /*tag=  ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                          module 2"
FT   misc_RNA        15763..16008
FT                   /*tag=  ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag=  ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag=  ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag=  af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag=  ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag=  ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag=  ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag=  aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag=  ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag=  al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag=  am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag=  an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag=  ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag=  ap
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag=  aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag=  ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag=  as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
```

```
FT                   /*tag=  at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag=  au
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag=  av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag=  aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag=  ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag=  ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag=  az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag=  ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag=  bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             98.2%;  Score 4187.6;  DB 21;  Length 71989;
  Best Local Similarity   98.9%;  Pred. No. 0;
  Matches 4217;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
Db       1998 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGGGC 2057

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
Db       2058 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 2117

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
Db       2118 CGCGACACCGTCGGGCAAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 2177

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
Db       2178 GACCTCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 2237

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
Db       2238 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 2297

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
Db       2298 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 2357

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
Db       2358 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 2417

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
Db       2418 GCGCTGCCGCGAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 2477

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
Db       2478 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGTCCTCGGGCTGCGAGGGCCGTGTGTCGCG 2537

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
Db       2538 GTGGATACGGCCTATTCGTCCTCGCTCGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 2597

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
Db       2598 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 2657

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
Db       2658 CTCGTGTGGCTCTCGAAGACCCGCGCGCTGGCCACGGACGGTCGCTGCAAGGCGTTTTCG 2717

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
Db       2718 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 2777

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
Db       2778 AGTGGAGCCCGCGCGGACGGCGACCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 2837

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
Db       2838 CACGACGGAGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 2897

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
Db       2898 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 2957

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
Db       2958 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 3017

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
Db       3018 CTCGGGCGAGACGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 3077

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
Db       3078 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 3137

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
Db       3138 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 3197

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
Db       3198 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 3257

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
Db       3258 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 3317

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
Db       3318 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGGCCGGCAGAGCTGCTGGTGCTGTCGGCA 3377

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
Db       3378 AGGACCGCGGCAGCCTTGGATGCACACGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 3437

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
Db       3438 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 3497

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
Db       3498 CACCGGCTCGCGGTGGCGGCGACGTCGAGCGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 3557

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
Db       3558 GCGCAGGGACAGACGCCGCCCGGTGTGGTGCGCGGTATCGCCGATTCCTCACGCGGCAAG 3617

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
Db       3618 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAT 3677

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
```

```
Db   3678 GATGTATGGCCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 3737

Qy   1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
Db   3738 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 3797

Qy   1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
Db   3798 CTCGACCAGACAGCCTTTACCCAGCCGGCGCTGTTCACCTTCGAGTATGCGCTCGCCGCG 3857

Qy   1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
Db   3858 CTGTGGCGGTCGTGGGGCGTAGAGCCGGAGTTGGTCGCTGGCCATAGCATCGGTGAGCTG 3917

Qy   1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
Db   3918 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 3977

Qy   1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
Db   3978 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGCGGCGCCG 4037

Qy   2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
Db   4038 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 4097

Qy   2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
Db   4098 AACGGTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 4157

Qy   2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
Db   4158 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 4217

Qy   2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
Db   4218 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 4277

Qy   2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
Db   4278 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTGGCACAGACGAGGTGAGC 4337

Qy   2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
Db   4338 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 4397

Qy   2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
Db   4398 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 4457

Qy   2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
Db   4458 GGCCTGGTGCCTGCCTGCCTGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 4517

Qy   2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
Db   4518 GGGCGTGACGAGCCAGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 4577

Qy   2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGTGCCGCTGCCCACGTAC 2640
Db   4578 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGTGCCGCTGCCCACGTAC 4637

Qy   2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
Db   4638 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 4697

Qy   2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
Db   4698 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGAAGGGGGGCGCGGTGCGCGGCGGC 4757

Qy   2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
Db   4758 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 4817

Qy   2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
Db   4818 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTCGATCGC 4877

Qy   2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
Db   4878 CTGGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTTGGCGAGGTCGAGATCGCCGTC 4937

Qy   2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
Db   4938 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 4997

Qy   3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
Db   4998 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 5057

Qy   3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
Db   5058 GCCGTGGGCGAGGGCGTGAACGGCCTTGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 5117

Qy   3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
Db   5118 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 5177

Qy   3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
Db   5178 CTCTCGGCGACCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 5237

Qy   3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
Db   5238 GACGGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCACGCGGCGACCGGCGGG 5297

Qy   3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
Db   5298 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 5357

Qy   3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
Db   5358 GGCACGCCCGAGAAGCGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 5417

Qy   3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
Db   5418 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 5477

Qy   3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
Db   5478 GTGCTCAACTCGCTTTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 5537

Qy   3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
Db   5538 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 5597

Qy   3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
Db   5598 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 5657

Qy   3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
Db   5658 CCGGCGCGGGTCCGTGCGCTCTTCGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 5717

Qy   3721 ACCCCTCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
Db   5718 ACCCCTCCCCCATCGCGACGCTCCCGATCGCTCGTGTCGCCGATGCGTTCCGGAGCATG 5777

Qy   3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
Db   5778 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 5837

Qy   3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
Db   5838 CGTATTCCGACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGATCTGCTCGACAGG 5897
```

```
Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5898 CTCGCGTCAGCTGCGCCGGCCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 5957

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5958 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 6017

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6018 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 6077

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       6078 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCAATATCGCCTTGTTGACCCAAAAC 6137

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6138 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 6197

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       6198 GCAGGCGTGCAAAGCGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 6257

Qy       4261 CTATGA 4266
              ||||||
Db       6258 CTATGA 6263
```

```
RESULT 3
ABQ77491
ID   ABQ77491 standard; DNA; 67459 BP.
XX
AC   ABQ77491;
XX
DT   14-MAY-2003  (first entry)
XX
DE   S. aurantiaca DNA containing sti gene cluster.
XX
KW   stiA; stiB; stiC; stiD; stiE; stiF; stiG; stiH; stiJ; stiK; stiL;
KW   aromatic structure biosynthesis; beta-ketoacylsynthase;
KW   acyltransferase; dehydratase; enoylreductase; beta-ketoacylreductase;
KW   acyl carrier protein; O-methyl transferase; ds.
XX
OS   Stigmatella aurantiaca.
XX
FH   Key             Location/Qualifiers
FT   CDS             14345..21466
FT                   /*tag=  a
FT                   /product= "stiA"
FT                   /note= "GTG start codon"
FT   CDS             21463..26220
FT                   /*tag=  b
FT                   /product= "stiB"
FT   CDS             26284..31941
FT                   /*tag=  c
FT                   /product= "stiC"
FT   CDS             31934..37738
FT                   /*tag=  d
FT                   /product= "stiD"
FT   CDS             37740..43553
FT                   /*tag=  e
FT                   /product= "stiE"
FT   CDS             43546..50202
FT                   /*tag=  f
FT                   /product= "stiF"
FT   CDS             50205..54401
FT                   /*tag=  g
FT                   /product= "stiG"
FT   CDS             54398..59212
FT                   /*tag=  h
FT                   /product= "stiH"
FT   CDS             59223..63002
FT                   /*tag=  i
FT                   /product= "stiJ"
FT   CDS             63070..63840
FT                   /*tag=  j
FT                   /product= "stiK"
FT   CDS             63929..65470
FT                   /*tag=  k
FT                   /product= "stiL"
FT                   /note= "GTG start codon"
XX
PN   DE10128661-A1.
XX
PD   19-DEC-2002.
XX
PF   15-JUN-2001; 2001DE-1028661.
XX
PR   15-JUN-2001; 2001DE-1028661.
XX
PA   (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX
PI   Kunze B,  Nordsiek G,  Silakowski B,  Bloecker H,  Gaitatzis N;
PI   Mueller R;
XX
DR   WPI; 2003-185404/19.
XX
PT   Nucleic acid molecule comprising one or more genes from Stigmatella
PT   species -
XX
PS   Claim 3; Page 9-43; 48pp; German.
XX
CC   This invention describes a novel nucleic acid molecule comprising one or
CC   more genes from Stigmatella species selected from stiA, stiB, stiC, stiD,
CC   stiE, stiF, stiG, stiH, stiJ stiK and stiL. The nucleic acid is used for
CC   biosynthetic purposes, particularly for producing aromatic structures.
CC   The products of the invention have domains sharing high homology with
CC   beta-ketoacylsynthase (KS), acyltransferase (AT), dehydratase (DH),
CC   enoylreductase (ER), beta-ketoacylreductase (KR), acyl carrier protein
CC   (ACP), O-methyl transferase (O-MT). This sequence represents a
CC   polynucleotide isolated from S. aurantiaca which contains the sti gene
CC   cluster described in the disclosure of the invention.
XX
SQ   Sequence 67459 BP; 10755 A; 20921 C; 23348 G; 12435 T; 0 other;

  Query Match             19.2%;  Score 817.2;  DB 25;  Length 67459;
  Best Local Similarity   57.5%;  Pred. No. 3.1e-122;
  Matches 1597;  Conservative    0;  Mismatches 1148;  Indels   31;  Gaps    6;

Qy         21 GCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGT 80
              |||    |||   || || || |||||||| ||    |    || ||     ||||||| ||
Db      43614 GCGTCTGGCCCCCGAGCCCATCGCGATCGTGGGCATGGCCTGTCGCTACCCCGGTGGGGT 43673

Qy         81 GATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGT 140
              || ||      |   |  |||    |  |   || ||     ||  |||||  |    |
Db      43674 GAAGGACTCCGACTCCTATTGGGATTTGTTGAAGCGCGGCGGAGATGCCGTCACGGAGAT 43733

Qy        141 CCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAA 200
               || || |    |||||||      |     |||   |  ||   ||  ||| || ||
Db      43734 TCCAGCAGCCCGCTGGGACATCAACGCCTTCTACGATCCGAACCAGGACAGCCCCGGCAA 43793

Qy        201 GACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTT 260
              |||| |        | |    ||||||   || || |    | |||||  ||   | | ||
Db      43794 GACGACGACGCGCTGGGGGGGCTTCCTGGATGATGTGGAGGGGTTCGATCCCCACCTGTT 43853

Qy        261 CGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGT 320
              |||||||||||||| ||||||  | |  ||||| ||      || || ||||| |||||||
Db      43854 CGGCATCTCGCCTCGGGAAGCGGCGAGCATGGATCCCCACCAGCGCCTCTTCCTGGAGGT 43913

Qy        321 GTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAAC 380
                  |||||||||||| |||||||| |                |     || | |||  |   | ||
Db      43914 CGCCTGGGAGGCGCTCGAGAACGCAGGCCAGCAAACCGATCGTCTGGCCGGGAGCAAGAC 43973

Qy        381 GGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGC 440
```

```
Db    43974 GGGGTCTTCGTCGGCCTTTACACGAGCGACTACTCTCCTGTTGTTC------TCTCGGA 44027

Qy      441 GTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCG 500
Db    44028 TCTCAACGCCATCAACGTCTATACCGGCACAGGCACGGCCAACAACGTGGCGGCTGGACG 44087

Qy      501 AATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTC 560
Db    44088 ATTGTCCTATCTGCTGGACCTGCAAGGGCCCAGCTTTGTGGTGGACACGGCCTGCTCCTC 44147

Qy      561 CTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGC 620
Db    44148 CTCGCTTGTCGCCGTTCACCTGGCGTGTCAGAGCCTTCTGAATCGTGAATGCAGAATGGC 44207

Qy      621 CCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGAC 680
Db    44208 GGTCGTGGGCGGCGTGACGTTGTTGCTCACGCCACTGACCCTCGTCATGGCGTCGAAGAT 44267

Qy      681 CCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGG 740
Db    44268 GCAGCTCATGGCCGCGGACGGCCGCTGCAAGACGTTCGACAGCCGCGCCGATGGTATCGC 44327

Qy      741 ACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGG 800
Db    44328 CATGGGTGAGGGCTGTGGCGTGGTGATCCTCAAGCGCTTGTCGGATGCCCAGGCCGATGG 44387

Qy      801 CGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGG 860
Db    44388 TGATCGCATCCTGGCGGTGATCAAGGCTTCGGCGATCAATCAAGACGGCCGCACGAATGG 44447

Qy      861 TCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGC 920
Db    44448 GCTGACAGCGCCGAGTGGCGCTTCCCAGCAGGTGTTGTTGCGTCAGGCGCTGACGCGCGC 44507

Qy      921 AGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGG 980
Db    44508 TGGCATCTCGCCGTCGCAGGTGACGTACGTGGAAGCCCACGGTACGGGGACACCGCTCGG 44567

Qy      981 TGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCAC 1040
Db    44568 CGATCCGATTGAGGTGGAGGCCCTCGCGGAGGTCATCGGGCAGCCACGACCTTCCGGGCC 44627

Qy     1041 GCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGAT 1100
Db    44628 TCCCTGTGTCCTGGGTTCCGTGAAGGCCAATCTGGGACATACCGGCGCGGCGGCGGGGGT 44687

Qy     1101 CACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCA 1160
Db    44688 CGCGGGGCTGATCAAAGTCGTGCTCTCCATGCAGAACGAGGCGATTCCTCCTCAACTGCA 44747

Qy     1161 CGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGC 1220
Db    44748 CTACCGGGAGCTCAACTCCAACATCTCCTTGGAGAACACCTCCCTGGTCATTCCGAAGGA 44807

Qy     1221 CCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCAT 1280
Db    44808 GCTGCATCCCTGGCCTGCGGGGAAGGGGCCGCGTATCGCAGGGATCAGTGCCTTCGGTTG 44867

Qy     1281 GAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCC 1340
Db    44868 GTCAGGGACCAACGCGCACGTGCTTGTGGAGGAGGCGCCTGCGGCCCCCAAGGCTCGCGC 44927

Qy     1341 GGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGGA 1400
Db    44928 GGCGGTTGAGCGTCCGGTGCACCTCCTGACCCTGTCGGCTCGGACCGAGGCCGCGCTGCG 44987

Qy     1401 TGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCGA 1460
Db    44988 CGAGTTCGCGGAGCGTGTGAAGACATATGTCTCGGCGCACGCGGATCTGCCCATCGCCGA 45047

Qy     1461 TGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGGC 1520
Db    45048 TCTCTGCTACTCGGCGAACACGGGGTGGTCGAAGTTCGAGCACCATCTGGCGGTCCCCGC 45107

Qy     1521 GACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGCC 1580
Db    45108 GGGGACGACGGCCCAGTTGAGCGAGCGTCTCTCCGCGTTTCTTGAGGGACAGCCGCAAGC 45167

Qy     1581 CGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG---CTCGCCTTTCTCTTCAC 1637
Db    45168 AGGCACGTCGTACGGACAGGTCACGGGACGCCGTGGGCCGGGAGTCGCGTTCCTGTTCAC 45227

Qy     1638 CGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGTT 1697
Db    45228 GGGGCAGGGCTCGCAGTACGCGGGCATGGGGCGGCAACTCTACCAGACACAGCCCACATT 45287

Qy     1698 CCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCCG 1757
Db    45288 CCGGCGTGCCCTCGATCGGTGCGACGAGTTGCTGCGCCCCTTGCTCGGGCGCTCCATTCT 45347

Qy     1758 CGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCTT 1817
Db    45348 GTCGGTGATCTACCCCGAAGCG------GGGCAGGCTTCTCCGATCGACGAGACGGCCTT 45401

Qy     1818 CACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGGG 1877
Db    45402 CACGCAGCCCGCGTTGTTCGCGTTCGAGTGGGCGCTCGCGGAACTCTGGCGCTCGTGGGG 45461

Qy     1878 TGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGGC 1937
Db    45462 CGTTGTACCGACCGTGGTGATGGGCCACAGCGTGGGTGAGTATGTGGCGGCCTGCGTCGC 45521

Qy     1938 GGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGCA 1997
Db    45522 TGGGGTGTTCACCCTCGAAGAAGGGCTTGCGCTCATCGTGGAGCGCGCCCGGTTGATGCA 45581

Qy     1998 GGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTGC 2057
Db    45582 GTCGTTGCCGCCGGGCGGCGAGATGGCCGCGGTCTTCACGGACGGGGCTCGTACCGCGCA 45641

Qy     2058 TGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCCGGACCAGGT 2117
Db    45642 GGCGATTGCTCCTTACGCATCGCAGGTCTCCATTGCCGCCTTCAATGGTCCTTCCGAGAC 45701

Qy     2118 GGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCGG 2177
Db    45702 GGTGATCTCTGGCGATGGCAAAGCCGTGGAGGCGATTCTCGGGGCGTTGTCCGCCGAGGG 45761

Qy     2178 GGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGAT 2237
Db    45762 CGTGAAGGTTCGGCGCCTGGGTGTCTCCCACGCCTTCCACTCTCCGCTCATGGACCCCAT 45821

Qy     2238 GCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCCT 2297
Db    45822 GCTGGATGCGTTCGAGCGGGCGGCGGCGAAGACGCGCTTCATGCCCCCGAAGATCACGCT 45881

Qy     2298 GGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATTGGGT 2357
Db    45882 GATCTCGAACCTCACCGG------TGGGCCCTCGGAGGAGTTCTCGGCCCGGGCGTTGCG 45935

Qy     2358 GCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGGCCGG 2417
Db    45936 TCGACATGCGCGTGAACCGGTCCGGTTCTCGGATGGGATGGCCGCGCTCAAGGCACGTGG 45995

Qy     2418 TGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCTGGTGCCTGCCTG 2477
Db    45996 CGTCTCGGTCGCGGTGGAGATTGGCCCGAATGCAACTCTGCTCGGCATCGGAAGCCGGTG 46055

Qy     2478 CATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGACGAGCCGGC 2537
Db    46056 TCTGCCGGAAGGGTCGGTGGCTTGGTTGCCTTCCCTTCGAAAGGACAAAGACGAGTGGGA 46115

Qy     2538 GACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCCTGGGCCGG 2597
Db    46116 AGTCCTGCTCGCAACGTTGGGTGCGCTCTCTGTGCGAGGCGTTCCCGTCAACTGGGCAGG 46175
```

```
Qy       2598 ---CCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGCAGCGCGA 2654
Db      46176 GTTCGACAGCGACTACCCGCGTCGGCGGGTCTCCCTGCCCAATTATCCGTTTCAGCACGA 46235

Qy       2655 GCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGACCGCCGTGCTCCGGG 2714
Db      46236 GCGGTTCGCGCTC-------AAGCAGATGGATGGACGGGGTGTGGCGCCGAGCGGTCAGG 46288

Qy       2715 AGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCGGCGCAGCGC 2774
Db      46289 TCGTGGGTCATCCCCTGTTGGGATCGCGCCTGCGCGTCGCGACGGTGGACGGCACTCTGT 46348

Qy       2775 TCGGCTCGACCATCCG 2790
Db      46349 TCGAATCGAAGTTCCG 46364
```

RESULT 4
AAC55785
ID   AAC55785 standard; DNA; 3978 BP.
XX
AC   AAC55785;
XX
DT   19-JAN-2001 (first entry)
XX
DE   Type I polyketide synthase orf8.
XX
KW   Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;
KW   anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   fungicide; pesticide; ds.
XX
OS   Streptomyces lavendulae.
XX
PN   WO200053737-A2.
XX
PD   14-SEP-2000.
XX
PF   10-MAR-2000; 2000WO-US06394.
XX
PR   12-MAR-1999;   99US-0266965.
XX
PA   (MINU) UNIV MINNESOTA.
PA   (SHER/) SHERMAN D H.
PA   (MAOY/) MAO Y.
PA   (VARO/) VAROGLU M.
PA   (HEMM/) HE M.
PA   (SHEL/) SHELDON P C.
XX
PI   Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX
DR   WPI; 2000-601980/57.
XX
PT   Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT   cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT   the molecular basis of mitosane ring system biosynthesis -
XX
PS   Example 1; Page 251-253; 399pp; English.
XX
CC   This invention relates to isolated and purified nucleic acid molecules
CC   from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC   natural products that contain a variety of functional groups, including
CC   amino benzoquinone and axiridine ring systems. The S. lavendulae
CC   mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC   spanning 55kb of DNA. The invention includes an expression cassette
CC   comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC   and host cells transformed with the cassette. The nucleotide, and protein
CC   sequences and the transformed host cells of the invention result in
CC   antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and
CC   antibiotic activities. The nucleotide sequences are used to elucidate the
CC   molecular basis for the biosynthesis of the mitosane ring system, as well
CC   as to engineer the biosynthesis of novel natural products, e.g.
CC   antibiotics, anti-inflammatory agents, anti-cancer agents,
CC   immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC   obstructive pulmonary disease as well as other disease involving
CC   respiratory inflammation, or cholesterol-lowering agents or as crop
CC   protection agents (e.g. fungicides or insecticides) as well as
CC   bipolymers, e.g., in packaging or biomedical applications, or to engineer
CC   PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and
CC   AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
CC   sequences and encoded proteins. Sequences AAC55812-C55814,
CC   AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
CC   cloning of the mitomycin biosynthetic genes.
XX
SQ   Sequence 3978 BP; 497 A; 1583 C; 1415 G; 483 T; 0 other;

Query Match 19.1%; Score 814.2; DB 21; Length 3978;
Best Local Similarity 57.6%; Pred. No. 1.1e-121;
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

```
Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77
Db         90 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 149

Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137
Db        150 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 209

Qy        138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194
Db        210 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 269

Qy        195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254
Db        270 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 329

Qy        255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314
Db        330 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 389

Qy        315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374
Db        390 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 449

Qy        375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434
Db        450 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 509

Qy        435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGC 494
Db        510 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 563

Qy        495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554
Db        564 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 623

Qy        555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614
Db        624 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 683

Qy        615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674
Db        684 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 743

Qy        675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734
Db        744 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 803

Qy        735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794
Db        804 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 863

Qy        795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854
Db        864 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 923
```

```
Qy     855 CAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914
Db     924 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 983

Qy     915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974
Db     984 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 1043

Qy     975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034
Db    1044 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGGCCCGCGGCGA 1103

Qy    1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094
Db    1104 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 1163

Qy    1095 GGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154
Db    1164 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 1223

Qy    1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214
Db    1224 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 1283

Qy    1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274
Db    1284 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 1343

Qy    1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334
Db    1344 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 1403

Qy    1335 ACCGCCG-----------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376
Db    1404 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 1463

Qy    1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436
Db    1464 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 1523

Qy    1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496
Db    1524 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 1583

Qy    1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556
Db    1584 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 1643

Qy    1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616
Db    1644 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 1703

Qy    1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGCCGTGGGCT 1676
Db    1704 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCCGGCATGGCCGCGAACT 1763

Qy    1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db    1764 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 1823

Qy    1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db    1824 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 1883

Qy    1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db    1884 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 1943

Qy    1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db    1944 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2003

Qy    1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
Db    2004 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 2063

Qy    1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db    2064 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 2123

Qy    2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db    2124 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 2183

Qy    2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
Db    2184 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 2243

Qy    2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216
Db    2244 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 2303

Qy    2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276
Db    2304 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 2363

Qy    2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336
Db    2364 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 2423

Qy    2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396
Db    2424 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 2483

Qy    2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456
Db    2484 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 2543

Qy    2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513
Db    2544 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 2603

Qy    2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGT 2573
Db    2604 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 2663

Qy    2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCC 2633
Db    2664 CGGCCTCTCGCCGGCGTGGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 2723

Qy    2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680
Db    2724 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 2770
```

RESULT 5
AAC55840
ID   AAC55840 standard; DNA; 12249 BP.
XX
AC   AAC55840;
XX
DT   19-JAN-2001 (first entry)
XX
DE   Complete Mitomycin ORF 1-9 nucleotide sequence.
XX
KW   Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;
KW   anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   fungicide; pesticide; ds.
XX
OS   Streptomyces lavendulae.
XX
PN   WO200053737-A2.
XX
PD   14-SEP-2000.

```
XX
PF   10-MAR-2000; 2000WO-US06394.
XX
PR   12-MAR-1999;   99US-0266965.
XX
PA   (MINU ) UNIV MINNESOTA.
PA   (SHER/) SHERMAN D H.
PA   (MAOY/) MAO Y.
PA   (VARO/) VAROGLU M.
PA   (HEMM/) HE M.
PA   (SHEL/) SHELDON P C.
XX
PI   Sherman DH,  Mao Y,  Varoglu M,  He M,  Sheldon PC;
XX
DR   WPI; 2000-601980/57.
XX
PT   Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT   cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT   the molecular basis of mitosane ring system biosynthesis  -
XX
PS   Claim 26; Figure 22; 399pp; English.
XX
CC   This invention relates to isolated and purified nucleic acid molecules
CC   from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC   natural products that contain a variety of functional groups, including
CC   amino benzoquinone and axiridine ring systems. The S. lavendulae
CC   mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC   spanning 55kb of DNA. The invention includes an expression cassette
CC   comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC   and host cells transformed with the cassette. The nucleotide, and protein
CC   sequences and the transformed host cells of the invention result in
CC   antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and
CC   antibiotic activities. The nucleotide sequences are used to elucidate the
CC   molecular basis for the biosynthesis of the mitosane ring system, as well
CC   as to engineer the biosynthesis of novel natural products, e.g.
CC   antibiotics, anti-inflammatory agents, anti-cancer agents,
CC   immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC   obstructive pulmonary disease as well as other disease involving
CC   respiratory inflammation, or cholesterol-lowering agents or as crop
CC   protection agents (e.g. fungicides or insecticides) as well as
CC   bipolymers, e.g., in packaging or biomedical applications, or to engineer
CC   PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and
CC   AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
CC   sequences and encoded proteins. Sequences AAC55812-C55814,
CC   AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
CC   cloning of the mitomycin biosynthetic genes.
XX
SQ   Sequence 12249 BP; 1715 A; 4722 C; 4196 G; 1616 T; 0 other;

  Query Match             19.1%;  Score 814.2;  DB 21;  Length 12249;
  Best Local Similarity   57.6%;  Pred. No. 1e-121;
  Matches 1549;  Conservative    0;  Mismatches 1108;  Indels   30;  Gaps    4;

Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77
Db       1077 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 1136

Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137
Db       1137 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 1196

Qy        138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194
Db       1197 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 1256

Qy        195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254
Db       1257 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 1316

Qy        255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314
Db       1317 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 1376

Qy        315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374
Db       1377 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 1436

Qy        375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434
Db       1437 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 1496

Qy        435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGC 494
Db       1497 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 1550

Qy        495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554
Db       1551 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 1610

Qy        555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614
Db       1611 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 1670

Qy        615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674
Db       1671 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 1730

Qy        675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734
Db       1731 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 1790

Qy        735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794
Db       1791 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 1850

Qy        795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854
Db       1851 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 1910

Qy        855 CAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914
Db       1911 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 1970

Qy        915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974
Db       1971 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 2030

Qy        975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034
Db       2031 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGCCCGCGGCGA 2090

Qy       1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094
Db       2091 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 2150

Qy       1095 GGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154
Db       2151 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 2210

Qy       1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214
Db       2211 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 2270

Qy       1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274
Db       2271 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 2330

Qy       1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334
Db       2331 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 2390

Qy       1335 ACCGCCG------------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376
Db       2391 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 2450

Qy       1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436
```

```
Db     2451 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 2510

Qy     1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496

Db     2511 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 2570

Qy     1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556

Db     2571 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 2630

Qy     1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616

Db     2631 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 2690

Qy     1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCT 1676

Db     2691 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCCGGCATGGGCCGCGAACT 2750

Qy     1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736

Db     2751 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 2810

Qy     1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796

Db     2811 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 2870

Qy     1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856

Db     2871 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 2930

Qy     1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916

Db     2931 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2990

Qy     1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976

Db     2991 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 3050

Qy     1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036

Db     3051 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 3110

Qy     2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096

Db     3111 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 3170

Qy     2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156

Db     3171 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 3230

Qy     2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216

Db     3231 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 3290

Qy     2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276

Db     3291 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 3350

Qy     2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336

Db     3351 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 3410

Qy     2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396

Db     3411 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 3470

Qy     2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456

Db     3471 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 3530

Qy     2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513

Db     3531 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 3590

Qy     2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGT 2573

Db     3591 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 3650

Qy     2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGCGGCGGGTGCCGCTGCC 2633

Db     3651 CGGCCTCTCGCCGGCGTGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 3710

Qy     2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680

Db     3711 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 3757
```

RESULT 6
AAC55857
ID   AAC55857 standard; DNA; 18331 BP.
XX
AC   AAC55857;
XX
DT   19-JAN-2001 (first entry)
XX
DE   Complete nucleotide sequence of the mitomycin gene cluster.
XX
KW   Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;
KW   anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   fungicide; pesticide; ds.
XX
OS   Streptomyces lavendulae.
XX
PN   WO200053737-A2.
XX
PD   14-SEP-2000.
XX
PF   10-MAR-2000; 2000WO-US06394.
XX
PR   12-MAR-1999; 99US-0266965.
XX
PA   (MINU) UNIV MINNESOTA.
PA   (SHER/) SHERMAN D H.
PA   (MAOY/) MAO Y.
PA   (VARO/) VAROGLU M.
PA   (HEMM/) HE M.
PA   (SHEL/) SHELDON P C.
XX
PI   Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX
DR   WPI; 2000-601980/57.
XX
PT   Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT   cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT   the molecular basis of mitosane ring system biosynthesis -
XX
PS   Disclosure; Figure 21; 399pp; English.
XX
CC   This invention relates to isolated and purified nucleic acid molecules
CC   from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC   natural products that contain a variety of functional groups, including
CC   amino benzoquinone and axiridine ring systems. The S. lavendulae
CC   mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC   spanning 55kb of DNA. The invention includes an expression cassette
CC   comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC   and host cells transformed with the cassette. The nucleotide, and protein
CC   sequences and the transformed host cells of the invention result in
CC   antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and
CC   antibiotic activities. The nucleotide sequences are used to elucidate the
CC   molecular basis for the biosynthesis of the mitosane ring system, as well
CC   as to engineer the biosynthesis of novel natural products, e.g.
CC   antibiotics, anti-inflammatory agents, anti-cancer agents,
CC   immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC   obstructive pulmonary disease as well as other disease involving

```
CC   respiratory inflammation, or cholesterol-lowering agents or as crop
CC   protection agents (e.g. fungicides or insecticides) as well as
CC   bipolymers, e.g., in packaging or biomedical applications, or to engineer
CC   PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and
CC   AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
CC   sequences and encoded proteins. Sequences AAC55812-C55814,
CC   AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
CC   cloning of the mitomycin biosynthetic genes.
XX
SQ   Sequence 18331 BP; 2523 A; 7003 C; 6343 G; 2462 T; 0 other;

  Query Match             19.1%;  Score 814.2;  DB 21;  Length 18331;
  Best Local Similarity   57.6%;  Pred. No. 1e-121;
  Matches 1549;  Conservative    0;  Mismatches 1108;  Indels   30;  Gaps    4;

Qy         18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77
Db       1077 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 1136

Qy         78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137
Db       1137 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 1196

Qy        138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194
Db       1197 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 1256

Qy        195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254
Db       1257 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 1316

Qy        255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314
Db       1317 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 1376

Qy        315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374
Db       1377 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 1436

Qy        375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434
Db       1437 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 1496

Qy        435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGC 494
Db       1497 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 1550

Qy        495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554
Db       1551 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 1610

Qy        555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614
Db       1611 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 1670

Qy        615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674
Db       1671 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 1730

Qy        675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734
Db       1731 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 1790

Qy        735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794
Db       1791 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 1850

Qy        795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854
Db       1851 CCTGGGCTACCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 1910

Qy        855 CAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914
Db       1911 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 1970

Qy        915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974
Db       1971 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 2030

Qy        975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034
Db       2031 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGCCCGCGGCGA 2090

Qy       1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094
Db       2091 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 2150

Qy       1095 GGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154
Db       2151 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 2210

Qy       1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214
Db       2211 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 2270

Qy       1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274
Db       2271 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 2330

Qy       1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334
Db       2331 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 2390

Qy       1335 ACCGCCG------------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376
Db       2391 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 2450

Qy       1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436
Db       2451 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 2510

Qy       1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496
Db       2511 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 2570

Qy       1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556
Db       2571 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 2630

Qy       1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616
Db       2631 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 2690

Qy       1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCT 1676
Db       2691 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCCGGCATGGGCCGCGAACT 2750

Qy       1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db       2751 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 2810

Qy       1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db       2811 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 2870

Qy       1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db       2871 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 2930

Qy       1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db       2931 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2990

Qy       1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
Db       2991 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 3050
```

```
Qy       1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db       3051 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 3110

Qy       2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db       3111 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 3170

Qy       2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
Db       3171 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 3230

Qy       2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216
Db       3231 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 3290

Qy       2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276
Db       3291 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 3350

Qy       2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336
Db       3351 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 3410

Qy       2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396
Db       3411 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 3470

Qy       2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456
Db       3471 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 3530

Qy       2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513
Db       3531 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCCCCCGAGACGTTCGTCCCCGTGCT 3590

Qy       2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGT 2573
Db       3591 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 3650

Qy       2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCC 2633
Db       3651 CGGCCTCTCGCCGGCGTGGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 3710

Qy       2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680
Db       3711 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 3757
```

RESULT 7
AAD17185
ID AAD17185 standard; DNA; 27541 BP.
XX
AC AAD17185;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; nys2; ds.
XX
OS Streptomyces noursei.
XX
FH Key Location/Qualifiers
FT CDS complement (454..1191)
FT /*tag= a
FT /product= "NysF protein"
FT /note= "CDS does not include start codon"
FT CDS complement (1275..3092)
FT /*tag= b
FT /product= "NysG protein"
FT CDS complement (3070..4824)
FT /*tag= c
FT /note= "CDS does not include start codon"
FT /product= "NysH protein"
FT CDS 5122..6156
FT /*tag= d
FT /product= "NysD3 protein"
FT CDS 6338..27541
FT /*tag= e
FT /product= "NysI partial protein"
FT /note= "CDS does not include stop codon"
XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.
XX
PR 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAERVIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
DR P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 2; Page 151-166; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

Query Match 18.9%; Score 805; DB 22; Length 27541;
Best Local Similarity 53.0%; Pred. No. 2.9e-120;
Matches 1996; Conservative 0; Mismatches 1715; Indels 52; Gaps 11;

```
Qy          4 GCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGC 63
Db      21008 GCCGTCCCGACCGGCGCCCCCGACGCCGACGACCCGATCGCGATCGTCGGCATGAGCTGC 21067

Qy         64 CGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGC 123
Db      21068 CGCTACCCCGGCGGGGTCGGCTCCGCCGAGGACCTGTGGCGGATCGCCCTGGACGAGGTC 21127

Qy        124 GACACCGTCGGGCGAGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCC 180
Db      21128 GACGCGATCTCCGGCTTCCCCGCCGACCGCGGCTGGGACGCCGAGGGCCTCTACGACCCG 21187

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
Db      21188 GACCCCGACCGGCCCGGCCGCACCTACTCCGTCCAGGGCGGATTCCTGCGCGACGTCGCC 21247

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
```

```
Db    21248 GAGTTCGACCCGGGCTTCTTCGGGATCTCGCCGCGCGAGGCGCTGTCGATGGACCCGCAG 21307

Qy      301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
Db    21308 CAGCGGCTCCTGCTGGAGACCGCCTGGGAGGCGTTCGAGCACGCCGGCATCGACCCGGTC 21367

Qy      361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
Db    21368 GGCCAGCGCGGCAGCCGCACCGGCACCTTCGTCGGCGCCAGCTACCAGGACTACGCCTCC 21427

Qy      421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
Db    21428 GGCGTGCCCAACAGCGAGGGCTC---------CGAAGGCCACATGATCACCGGCACGCTC 21478

Qy      481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
Db    21479 TCCAGTGTGCTGTCCGGCCGGGTGTCCTACCTCTTCGGCTTCGAGGGCCCCGCCGTCACG 21538

Qy      541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
Db    21539 CTCGACACCGCCTGCTCCTCCTCCCTGGTCGCCATGCACCTGGCCTGCCAGTCCCTGCGC 21598

Qy      601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
Db    21599 AACGGGGAGAGCTCGCTGGCCCTGGCCGGCGGCGTCAGCATCATGTCCACCCCGATGTCG 21658

Qy      661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
Db    21659 TTCGTCGGCTTCAGCCGGCAGCGCGCCCTCGCCGAGGACGGCCGCTGCAAGGCGTACGCG 21718

Qy      721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
Db    21719 GACGGCGCCGACGGGATGACCCTCGCCGAGGGCGTCGGCCTGGTGCTGCTGGAGCGGCTG 21778

Qy      781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
Db    21779 TCCGACGCCCGCGCCAACGGGCACCAGGTGCTCGCCGTGATCCGCGGCTCCGCCGTCAAC 21838

Qy      841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
Db    21839 CAGGACGGCGCCTCCAACGGCCTGACCGCACCCAACGGCCCGTCCCAGCAGCGCGTCATC 21898

Qy      901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
Db    21899 CGCCAGGCGCTGGCCAACTCCGCGGTGGCGCCCGGCGACATCGACGTCCTGGAGGGCCAC 21958

Qy      961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
Db    21959 GGCACCGGTACCGCCCTCGGCGACCCCATCGAGGCGCAGGCCCTGCTCGCCACCTACGGC 22018

Qy     1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
Db    22019 CAGGACCGCGCCCCCGAACGGCCGCTGCTGCTCGGCTCGGTGAAGTCCAACATCGGCCAC 22078

Qy     1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
Db    22079 ACCCAGATGGCATCCGGCGTCGCCAGCGTCATCAAGCTCGTCCGCGCCCTCCAGGAAGGC 22138

Qy     1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
Db    22139 GTGGTGCCCAAGTCCCTGCACATCGACCGGCCCTCCACCCACGTCGACTGGTCCTCGGGC 22198

Qy     1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
Db    22199 GCCATCGGGCTGCTCACCGAACGCACCCCGTGGCCCGAGACCGGCCGGCCGCGCCGCGCC 22258

Qy     1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
Db    22259 GCCGTCTCCTCCTTCGGCATCAGCGGCACCAACGTCCACACCATCCTCGAACAGGCCCCC 22318

Qy     1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTG------GTGCTG 1374
Db    22319 GCGGACGAGGCGCCCACGCCCGCCGACCCGCCGCGGGACGGCCTGGTGCCGGTCCTGCTC 22378

Qy     1375 TCGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAG 1434
Db    22379 TCCGGCCGCGGCGAGGCCGCGCTGCGCGCCCAGGCCGCCCGCCTGCTCGCCTTCGTCGAG 22438

Qy     1435 ACCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCG 1494
Db    22439 GAGCGGCCCGAGGCCCACCTCACCGACCTCGCCCACTCCCTCGCCACCTCGCGCGCCGCG 22498

Qy     1495 ATGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGAC 1554
Db    22499 CTGGAACGCCGCGCCGCCGTCATCGCCGCCGACCGCGACACCCTGACCCGCGGCCTGCGC 22558

Qy     1555 GCTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGC 1614
Db    22559 GCCCTGTCCGACGGCCGGCCCGACCCGGCCTGGTCCAGGGCACCGCGG------ACGC 22612

Qy     1615 GGCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGG 1674
Db    22613 GGCCGGACCGCCTTCCTGTTCACCGGACAGGGCAGCCAGCGCCCCGGCATGGGCCGCGAA 22672

Qy     1675 CTGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAAC 1734
Db    22673 CTCCACGACCGCTACCCGGTGTTCGCCGACGCGCTGGACGAGGTGCTGGCCCGGCTCGAC 22732

Qy     1735 CAGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCC 1794
Db    22733 GACGGACCGGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGCCCGACTCCGCCGAGGCC 22792

Qy     1795 GCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTC 1854
Db    22793 GCGCTCCTGGACCGGACCGGCTACGCCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTG 22852

Qy     1855 GCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGT 1914
Db    22853 TTCCGCCTGCTGACGTCCTGGGGCCTGACCCCGGACTACCTGGCCGGCCACTCCGTCGGC 22912

Qy     1915 GAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTG 1974
Db    22913 GAACTCGCCGCCGCGCACGTCGCCGGCGTGCTGTCGCTGGACGACGCCTGCACTCTGGTC 22972

Qy     1975 GCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAG 2034
Db    22973 GCCGCCCGCGGCCGGCTCATGCAGGCGCTGCCCGAGGGCGGCGCGATGGTCGCCCTGGAG 23032

Qy     2035 GCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCC 2094
Db    23033 GCCGCGGAGGACGAGGTCCTGCCGCTCCTGGAGGGGCTCACCGACCGGGTGTCCGTCGCC 23092

Qy     2095 GCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATC 2154
Db    23093 GCCGTCAACGGGCCGCGGTCCGTGGTCGTCGCCGGCGTCGAGGAGGACGTGCTCCTCCTC 23152

Qy     2155 GCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTC 2214
Db    23153 GCCGACCTCTTCGCCGCCGACGGGCGCCGGACCAAGCGGCTGCGGGTGAGCCATGCCTTC 23212

Qy     2215 CACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGC 2274
Db    23213 CACTCGCCGCTGATGGACGCCATGCTCGACGACTTCGCCGCCGTCGCCCGCGGGCTGACC 23272

Qy     2275 TACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGA---GCGGGAAGGCTTGCACAGAC 2331
Db    23273 TACCACCCGCCGACGATCCCGTTCGTGTCGAACGTCAGCGGCGGCCTGGCCACCGCCGAA 23332

Qy     2332 GAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGAT 2391
Db    23333 CAGGTCCGCACGCCCGACTACTGGGTCGGGCACGTCCGCGCCGCGGTCCGCTTCGCCGAC 23392

Qy     2392 GGAGTGAAGGCGCTGCACGC---GGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAA 2448
Db    23393 GGCATCGACTGGCTCGCCACCCAGGGCGACGTCCACACCTTCCTGGAGCTCGGCCCCGGAC 23452
```

```
Qy       2449 TCGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACG---CCCGGCCGGCGCTGCTC 2505
Db      23453 GGCGTGCTCAGCGCCATGGCCCGGGAGAGCCTCACCGACCCGTCCCGCACGGCACTGCTG 23512

Qy       2506 GCATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTC 2565
Db      23513 CCGACCCTGCGCGGCGACCGGCCCGAGGAACCTGCCCTGGTCACCGCGGTCGCCGCGGCC 23572

Qy       2566 TGGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCC---CTCAGGGGGGCGGCGG 2622
Db      23573 CACGCGCACGGCGCCCGCGTCGACTGGAGCGGGTACTTCGCCGACCACGGCGCGCGCCGG 23632

Qy       2623 GTGCCGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGAC 2682
Db      23633 ACCACGCTGCCGACCTACGCGTTCCAACGCGAGCGGTACTGGCCCGACACCACCGCCGCC 23692

Qy       2683 GACGCGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGG 2742
Db      23693 ACGAGCGCCCACACGCCCGGATCCGCCCTCGACGCCGAGTTCTGGGCCGCCGTCGAGCGG 23752

Qy       2743 GGCGCGGTGCGCGGCGGCGACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGC 2802
Db      23753 GACGACGTCGCCGCCCTCGCCGCCTCCCTGGACCTGGACGACGCCACCGTCACCGCGATG 23812

Qy       2803 GGACGCCGGGAGAAGGTCGAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAG 2862
Db      23813 GTCCCCGCGCTCACCGCCTGGCGCCGGCGCCGCGGCGAG---CAGACCGAACTGGACTCC 23869

Qy       2863 CCAGGCGTGCTTGATCACCTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGC 2922
Db      23870 TGGCGCTACCGCGTCACCTGGAAGCCGCGCGGCGGCGCCACCGCACCCGCCGCCCTCACC 23929

Qy       2923 GAGGTCGAGATCGCCGTCGACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTG 2982
Db      23930 GGCCGCTGGCTCGTGCTCGTCCCGCACGACCACCAGGACCGTCAGGACGACGCGACCGCG 23989

Qy       2983 GGCATGGTGCCCGACGACCTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAG 3042
Db      23990 GCCTGGGCAGCCGACGTCGAGACCGCCCTGGGCACCACCACCGTCCGGCTGACGGTCACC 24049

Qy       3043 TGCGCCGGGCGCATCGTCGCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCG 3102
Db      24050 ACCACCGACCGCGCCGCGCTGGCCGCCCGGATCACCGAAGCCGCCG-----------GCG 24098

Qy       3103 GTCATCGCCCTTTCGGCGGGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTG 3162
Db      24099 ACCAGGGCCCGTTCAGCGGTGTGCTGTCCCTGCTGCCGCTCGCCACCGGCGACGCCGGCC 24158

Qy       3163 CTGCCTCGGCCTCAGGCGCTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTG 3222
Db      24159 ACCCCGGTGCGCCCGCCGCCCTCACCCTCACCACCACCGCCGTCCAGGCCCTCGGCGACG 24218

Qy       3223 ACGGCATGGTACGCGCTCGACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATC 3282
Db      24219 CCGGCATCGACGCGCCGCTGTGGAACGTCACCCGCGGGGCCGTGGCCGTCGGCCGCGCCG 24278

Qy       3283 CATGCGGCGACCGGCGGGGTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCC 3342
Db      24279 AACAGGTCACCGCGCCCGAACAGGCCGCCGTCTGGGGCCTGGGCCGCGCCGTCGCCCTGG 24338

Qy       3343 GAGGTCCATGC--GACGGCCGGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCG 3400
Db      24339 AACTGCCGGCCCGGTTCGGCGGCACCCTCGACCTGCCCGCCACCCTGGACGGCCAGGCCG 24398

Qy       3401 TGCGGTATGTGAGCGATTCCCGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGG 3460
Db      24399 CCCGCCGGTTGCGCGCGGTGCTCGCGGCTACCGACGGCGAGGACGCGGTGGCCCTGCGGC 24458

Qy       3461 GCGGCGAGGGAGTAGACGTCGTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTT 3520
Db      24459 CCTCCGGCGTCTTCCTCCGCCGCCTGGCCCACGCCCCGGCCGGCCCCGACACCGCCCGCA 24518

Qy       3521 TCAATCTCCTGCGATCGCACGGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGG 3580
Db      24519 CCGCCTTCGACCCGGCGGCCGGCACCGTCCTGATCACCGGCGGCACCGGCGGCATCGGCG 24578

Qy       3581 ATAACCAGCTCGGGCTGCGGCCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCC 3640
Db      24579 GCCACGTCGCCCGCCGCCTGGCCCGCGACGGCGCCACCCACCTGCTGCTCACCAGCCGCC 24638

Qy       3641 GGGGGATGATGCTCGAGCGGCCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCC 3700
Db      24639 GCGGCCCGGCCGCCCCCGGCGCCGACGCGCTCCGCGCCGAACTGGAGGAACTGGGCGCCC 24698

Qy       3701 TGATCGCGGCAGGCGTGTTCACCCCTCCCCCATCGCGACGCT 3743
Db      24699 GGGTCACCCTCGCCGCCTGCGACGCCGCCGACCGCGACGCGCT 24741
```

```
RESULT 8
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag=  a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag=  b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag=  c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag=  d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag=  e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag=  f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag=  g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag=  h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
```

```
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             18.9%;  Score 805;  DB 22;  Length 125401;
  Best Local Similarity   53.0%;  Pred. No. 2.7e-120;
  Matches 1996;  Conservative    0;  Mismatches 1715;  Indels   52;  Gaps   11;

Qy          4 GCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGC 63
Db      21008 GCCGTCCCGACCGGCGCCCCCGACGCCGACGACCCGATCGCGATCGTCGGCATGAGCTGC 21067

Qy         64 CGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGC 123
Db      21068 CGCTACCCCGGCGGGGTCGGCTCCGCCGAGGACCTGTGGCGGATCGCCCTGGACGAGGTC 21127

Qy        124 GACACCGTCGGGCGAGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCC 180
Db      21128 GACGCGATCTCCGGCTTCCCCGCCGACCGCGGCTGGGACGCCGAGGGCCTCTACGACCCG 21187

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
Db      21188 GACCCCGACCGGCCCGGCCGCACCTACTCCGTCCAGGGCGGATTCCTGCGCGACGTCGCC 21247

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
Db      21248 GAGTTCGACCCGGGCTTCTTCGGGATCTCGCCGCGCGAGGCGCTGTCGATGGACCCGCAG 21307

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
Db      21308 CAGCGGCTCCTGCTGGAGACCGCCTGGGAGGCGTTCGAGCACGCCGGCATCGACCCGGTC 21367

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
Db      21368 GGCCAGCGCGGCAGCCGCACCGGCACCTTCGTCGGCGCCAGCTACCAGGACTACGCCTCC 21427

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
Db      21428 GGCGTGCCCAACAGCGAGGGCTC---------CGAAGGCCACATGATCACCGGCACGCTC 21478

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
Db      21479 TCCAGTGTGCTGTCCGGCCGGGTGTCCTACCTCTTCGGCTTCGAGGGCCCCGCCGTCACG 21538

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
Db      21539 CTCGACACCGCCTGCTCCTCCTCCCTGGTCGCCATGCACCTGGCCTGCCAGTCCCTGCGC 21598

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
Db      21599 AACGGGGAGAGCTCGCTGGCCCTGGCCGGCGGCGTCAGCATCATGTCCACCCCGATGTCG 21658

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
Db      21659 TTCGTCGGCTTCAGCCGGCAGCGCGCCCTCGCCGAGGACGGCCGCTGCAAGGCGTACGCG 21718

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
Db      21719 GACGGCGCCGACGGGATGACCCTCGCCGAGGGCGTCGGCCTGGTGCTGCTGGAGCGGCTG 21778

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
Db      21779 TCCGACGCCCGCGCCAACGGGCACCAGGTGCTCGCCGTGATCCGCGGCTCCGCCGTCAAC 21838

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
Db      21839 CAGGACGGCGCCTCCAACGGCCTGACCGCACCCAACGGCCCGTCCCAGCAGCGCGTCATC 21898

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
Db      21899 CGCCAGGCGCTGGCCAACTCCGCGGTGGCGCCCGGCGACATCGACGTCCTGGAGGGCCAC 21958

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
Db      21959 GGCACCGGTACCGCCCTCGGCGACCCCATCGAGGCGCAGGCCCTGCTCGCCACCTACGGC 22018

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
Db      22019 CAGGACCGCGCCCCCGAACGGCCGCTGCTGCTCGGCTCGGTGAAGTCCAACATCGGCCAC 22078

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
Db      22079 ACCCAGATGGCATCCGGCGTCGCCAGCGTCATCAAGCTCGTCCGCGCCCTCCAGGAAGGC 22138

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
Db      22139 GTGGTGCCCAAGTCCCTGCACATCGACCGGCCCTCCACCCACGTCGACTGGTCCTCGGGC 22198

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
Db      22199 GCCATCGGGCTGCTCACCGAACGCACCCCGTGGCCCGAGACCGGCCGGCCGCGCCGCGCC 22258

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
Db      22259 GCCGTCTCCTCCTTCGGCATCAGCGGCACCAACGTCCACACCATCCTCGAACAGGCCCCC 22318

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTG------GTGCTG 1374
Db      22319 GCGGACGAGGCGCCCACGCCCGCCGACCCGCCGCGGGACGGCCTGGTGCCGGTCCTGCTC 22378

Qy       1375 TCGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAG 1434
Db      22379 TCCGGCCGCGGCGAGGCCGCGCTGCGCGCCCAGGCCGCCCGCCTGCTCGCCTTCGTCGAG 22438

Qy       1435 ACCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCG 1494
Db      22439 GAGCGGCCCGAGGCCCACCTCACCGACCTCGCCCACTCCCTCGCCACCTCGCGCGCCGCG 22498

Qy       1495 ATGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGAC 1554
Db      22499 CTGGAACGCCGCGCCGCCGTCATCGCCGCCGACCGCGACACCCTGACCCGCGGCCTGCGC 22558

Qy       1555 GCTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGC 1614
Db      22559 GCCCTGTCCGACGGCCGGCCCGACCCCGGCCTGGTCCAGGGCACCGCGGG------ACGC 22612

Qy       1615 GGCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGG 1674
Db      22613 GGCCGGACCGCCTTCCTGTTCACCGGACAGGGCAGCCAGCGCCCCGGCATGGGCCGCGAA 22672

Qy       1675 CTGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAAC 1734
```

```
Db   22673 CTCCACGACCGCTACCCGGTGTTCGCCGACGCGCTGGACGAGGTGCTGGCCCGGCTCGAC 22732

Qy    1735 CAGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCC 1794

Db   22733 GACGGACCGGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGCCCGACTCCGCCGAGGCC 22792

Qy    1795 GCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTC 1854

Db   22793 GCGCTCCTGGACCGGACCGGCTACGCCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTG 22852

Qy    1855 GCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGT 1914

Db   22853 TTCCGCCTGCTGACGTCCTGGGGCCTGACCCCGGACTACCTGGCCGGCCACTCCGTCGGC 22912

Qy    1915 GAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTG 1974

Db   22913 GAACTCGCCGCCGCGCACGTCGCCGGCGTGCTGTCGCTGGACGACGCCTGCACTCTGGTC 22972

Qy    1975 GCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAG 2034

Db   22973 GCCGCCCGCGGCCGGCTCATGCAGGCGCTGCCCGAGGGCGGCGCGATGGTCGCCCTGGAG 23032

Qy    2035 GCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCC 2094

Db   23033 GCCGCGGAGGACGAGGTCCTGCCGCTCCTGGAGGGGCTCACCGACCGGGTGTCCGTCGCC 23092

Qy    2095 GCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATC 2154

Db   23093 GCCGTCAACGGGCCGCGGTCCGTGGTCGTCGCCGGCGTCGAGGAGGACGTGCTCCTCCTC 23152

Qy    2155 GCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTC 2214

Db   23153 GCCGACCTCTTCGCCGCCGACGGGCGCCGGACCAAGCGGCTGCGGGTGAGCCATGCCTTC 23212

Qy    2215 CACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGC 2274

Db   23213 CACTCGCCGCTGATGGACGCCATGCTCGACGACTTCGCCGCCGTCGCCCGCGGGCTGACC 23272

Qy    2275 TACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGA---GCGGGAAGGCTTGCACAGAC 2331

Db   23273 TACCACCCGCCGACGATCCCGTTCGTGTCGAACGTCAGCGGCGGCCTGGCCACCGCCGAA 23332

Qy    2332 GAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGAT 2391

Db   23333 CAGGTCCGCACGCCCGACTACTGGGTCGGGCACGTCCGCGCCGCGGTCCGCTTCGCCGAC 23392

Qy    2392 GGAGTGAAGGCGCTGCACGC---GGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAA 2448

Db   23393 GGCATCGACTGGCTCGCCACCCAGGGCGACGTCCACACCTTCCTGGAGCTCGGCCCGGAC 23452

Qy    2449 TCGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACG---CCCGGCCGGCGCTGCTC 2505

Db   23453 GGCGTGCTCAGCGCCATGGCCCGGGAGAGCCTCACCGACCCGTCCCGCACGGCACTGCTG 23512

Qy    2506 GCATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTC 2565

Db   23513 CCGACCCTGCGCGGCGACCGGCCCGAGGAACCTGCCCTGGTCACCGCGGTCGCCGCGGCC 23572

Qy    2566 TGGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCC---CTCAGGGGGGCGGCGG 2622

Db   23573 CACGCGCACGGCGCCCGCGTCGACTGGAGCGGGTACTTCGCCGACCACGGCGCGCGCCGG 23632

Qy    2623 GTGCCGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGAC 2682

Db   23633 ACCACGCTGCCGACCTACGCGTTCCAACGCGAGCGGTACTGGCCGACACCACCGCCGCC 23692

Qy    2683 GACGCGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGG 2742

Db   23693 ACGAGCGCCCACACGCCCGGATCCGCCCTCGACGCCGAGTTCTGGGCCGCCGTCGAGCGG 23752

Qy    2743 GGCGCGGTGCGCGGCGGCGACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGC 2802

Db   23753 GACGACGTCGCCGCCCTCGCCGCCTCCCTGGACCTGGACGACGCCACCGTCACCGCGATG 23812

Qy    2803 GGACGCCGGGAGAAGGTCGAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAG 2862

Db   23813 GTCCCCGCGCTCACCGCCTGGCGCCGGCGCCGCGGCGAG---CAGACCGAACTGGACTCC 23869

Qy    2863 CCAGGCGTGCTTGATCACCTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGC 2922

Db   23870 TGGCGCTACCGCGTCACCTGGAAGCCGCGCGGCGGCGCCACCGCACCCGCCGCCCTCACC 23929

Qy    2923 GAGGTCGAGATCGCCGTCGACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTG 2982

Db   23930 GGCCGCTGGCTCGTGCTCGTCCCGCACGACCACCAGGACCGTCAGGACGACGCGACCGCG 23989

Qy    2983 GGCATGGTGCCCGACGACCTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAG 3042

Db   23990 GCCTGGGCAGCCGACGTCGAGACCGCCCTGGGCACCACCACCGTCCGGCTGACGGTCACC 24049

Qy    3043 TGCGCCGGGCGCATCGTCGCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCG 3102

Db   24050 ACCACCGACCGCGCCGCGCTGGCCGCCCGGATCACCGAAGCCGCCG----------GCG 24098

Qy    3103 GTCATCGCCCTTTCGGCGGGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTG 3162

Db   24099 ACCAGGGCCCGTTCAGCGGTGTGCTGTCCCTGCTGCCGCTCGCCACCGGCGACGCCGGCC 24158

Qy    3163 CTGCCTCGGCCTCAGGCGCTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTG 3222

Db   24159 ACCCCGGTGCGCCCGCCGCCCTCACCCTCACCACCACCGCCGTCCAGGCCCTCGGCGACG 24218

Qy    3223 ACGGCATGGTACGCGCTCGACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATC 3282

Db   24219 CCGGCATCGACGCGCCGCTGTGGAACGTCACCCGCGGGGCCGTGGCCGTCGGCCGCGCCG 24278

Qy    3283 CATGCGGCGACCGGCGGGGTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCC 3342

Db   24279 AACAGGTCACCGCGCCCGAACAGGCCGCCGTCTGGGGCCTGGGCCGCGCCGTCGCCCTGG 24338

Qy    3343 GAGGTCCATGC--GACGGCCGGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCG 3400

Db   24339 AACTGCCGGCCCGGTTCGGCGGCACCCTCGACCTGCCCGCCACCCTGGACGGCCAGGCCG 24398

Qy    3401 TGCGGTATGTGAGCGATTCCCGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGG 3460

Db   24399 CCCGCCGGTTGCGCGCGGTGCTCGCGGCTACCGACGGCGAGGACGCGGTGGCCCTGCGGC 24458

Qy    3461 GCGGCGAGGGAGTAGACGTCGTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTT 3520

Db   24459 CCTCCGGCGTCTTCCTCCGCCGCCTGGCCCACGCCCCGGCCGGCCCCGACACCGCCCGCA 24518

Qy    3521 TCAATCTCCTGCGATCGCACGGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGG 3580

Db   24519 CCGCCTTCGACCCGGCGGCCGGCACCGTCCTGATCACCGGCGGCACCGGCGGCATCGGCG 24578

Qy    3581 ATAACCAGCTCGGGCTGCGGCCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCC 3640

Db   24579 GCCACGTCGCCCGCCGCCTGGCCCGCGACGGCGCCACCCACCTGCTGCTCACCAGCCGCC 24638

Qy    3641 GGGGGATGATGCTCGAGCGGCCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCC 3700

Db   24639 GCGGCCCGGCCGCCCCCGGCGCCGACGCGCTCCGCGCCGAACTGGAGGAACTGGGCGCCC 24698

Qy    3701 TGATCGCGGCAGGCGTGTTCACCCCTCCCCCCATCGCGACGCT 3743

Db   24699 GGGTCACCCTCGCCGCCTGCGACGCCGCCGACCGCGACGCGCT 24741
```

```
RESULT 9
AAS17367
ID   AAS17367 standard; DNA; 33529 BP.
XX
AC   AAS17367;
XX
DT   12-MAR-2002  (first entry)
```

```
XX
DE   DNA sequence of S. cellulosum polyketide synthase cosmid, pKOS28-26.
XX
KW   Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW   acyl transferase domain; dehydratase domain; ketoreductase domain;
KW   acyl carrier protein domain; pKOS28-26; ds.
XX
OS   Sorangium cellulosum.
XX
PN   US6280999-B1.
XX
PD   28-AUG-2001.
XX
PF   31-AUG-1998;   98US-0144085.
XX
PR   22-JAN-1998;   98US-0010809.
XX
PA   (KOSA-) KOSAN BIOSCIENCE.
XX
PI   Gustafsson C,  Betlach MC,  Ashley G,  Julien B,  Ziermann R;
XX
DR   WPI; 2001-606536/69.
XX
PT   Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT   polyketide open reading frame encoding modules with one or more domains
PT   such as ketosynthase, acyl transferase and acyl carrier protein domains
PT   -
XX
PS   Claim 4; Fig 1; 72pp; English.
XX
CC   The present invention relates to the isolation of novel Sorangium
CC   cellulosum polyketide synthases (PKS), and the polynucleotide sequences
CC   encoding them. The polyketide synthases include catalytic domains such
CC   as ketosynthase domain, acyl transferase domain, dehydratase domain,
CC   ketoreductase domain and acyl carrier protein domain. A host cell
CC   comprising a PKS ORF (open reading frame) which encodes one or more
CC   more PKS domains is useful for producing polyketide synthases from which
CC   polyketides can be produced. The host cells are useful for constructing
CC   a library, where each individual colony of the library represents a
CC   colony with the ability to produce a particular PKS synthase and
CC   ultimately a particular polyketide. The polyketides produced by these
CC   colonies can be used collectively in a panel to represent a library or
CC   may be assessed individually for activity. Colonies in the library are
CC   also induced to produce the relevant synthases and thus to produce the
CC   relevant polyketides to obtain a library of candidate polyketides which
CC   can be screened for binding to desired targets such as receptors,
CC   signalling proteins, etc. The present sequence represents the DNA
CC   sequence of cosmid pKOS28-26 which encodes one or more domains of
CC   S. cellulosum PKS.
CC   Note: The present sequence is said to encode the functional domains
CC   of S. cellulosum PKS which correspond to domains or domain subsets of
CC   the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX
SQ   Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other;

  Query Match             18.8%;  Score 803.2;  DB 23;  Length 33529;
  Best Local Similarity   57.6%;  Pred. No. 5.6e-120;
  Matches 1591;  Conservative    0;  Mismatches 1128;  Indels   41;  Gaps    7;

Qy         37 CCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGG 96
Db      25316 CCGATAGCGATCGTGGGCGTGGGGCTGCGCATGCCTGGCGGAGCGAGCGACCTGGAGAGC 25375

Qy         97 TTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGG 156
Db      25376 TTCTGGCAGGTGCTGGTCGAAGAGCGGGATACGCTGCGGCCGATCCCGGCCCAACGATTC 25435

Qy        157 GATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGC 216
Db      25436 GACGTCGAGGCGCTGTACGATCCTGACCCCGACGCGAAGGGCAAGACGTACGTGCGGAAC 25495

Qy        217 GCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGC 276
Db      25496 GCGTCGCTGCTCGACGACGTGGCGTCGTTCGACCCTGGGTTCTTCGGGATAAGTCCGCGG 25555

Qy        277 GAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTG 336
Db      25556 GAGGCGGAGCCGATGGATCCGCAGCACCGGCTGCTGCTGGAGACGGCGTGGAGCGCCCTG 25615

Qy        337 GAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGG 396
Db      25616 GAGGACGCGGGGGTGCGTCCAGAGCACCTGAAGGGCTCGGACACGGGAGTGTTCGTGGGC 25675

Qy        397 ATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGAC 456
Db      25676 GTGGCGCCGAGCGAATACGCGAGCTACCGAGGAAA---------GAGCGCGAACGAAGAT 25726

Qy        457 GCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTC 516
Db      25727 GCGTATGCGCTGACAGGGACGGCGCTGAGCTTTGCTGCGGGACGGGTGGCCTACCACCTC 25786

Qy        517 GGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTT 576
Db      25787 GGGCTGCAAGGCCCTGCGGTGTCGACCGACACGGCCTGCAGCTCGTCGCTGGTAGCGGTG 25846

Qy        577 CATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTA 636
Db      25847 CACCTGGCGTGCGACGCGCTGCGCCGGGGCGATTGCGAGGTGGCGCTGGCGGCAGGTGTG 25906

Qy        637 TCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGG 696
Db      25907 CAGGTGCTGGCGAACCCGGCGGGGTTTGTGCTGCTGTCGCGCACGCGCGCGTTGTCGCCG 25966

Qy        697 GACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGC 756
Db      25967 GACGGGCGGTGCAAGGCGTTCTCGCAGGCGGCGGACGGTTATGGCCGTGGCGAGGGAGTC 26026

Qy        757 GCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCG 816
Db      26027 GGGGTGCTGGTGCTGATGCGGCTGTCCGAGGCGCAGCAGCAGGGGAAGCGGGTGCTGGGT 26086

Qy        817 GTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAAC 876
Db      26087 GTGGTGCGCGGCACGGCGGTCAATCAGGACGGCGCGAGCAGCGGGATCACGGCGCCGAAC 26146

Qy        877 GGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCT 936
Db      26147 GGCACGGCGCAGCAGAAGGTGGTGCGCGCGGCGCTGCGGAACGCGGGGCTGGAGCCGGCG 26206

Qy        937 TCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGACCCCATCGAAATC 996
Db      26207 AGCATCGATGTGGTGGAGTGCCACGGTACGGGCACGTCGCTGGGCGACCCGATCGAGGTG 26266

Qy        997 CAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGG 1056
Db      26267 CAGGCGCTCGGCGCGGTGTACGGGCAAGGTCGGGATATGGCTCGTCCGCTGCAGCTGGGC 26326

Qy       1057 TCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAG 1116
Db      26327 GCGGTCAAGAGCAATATCGGTCATCTCGAGTCCGCCGCGGGCATCGCAGGGGTGTGCAAG 26386

Qy       1117 GTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAAC 1176
Db      26387 ATCCTGGCGGCGTTCCGTTACGAGTCGCTGCCGGCGACGCTGCACAGCTCGCCGCGCAAT 26446

Qy       1177 CCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCGGACACCGTGGCC- 1235
Db      26447 CCCCGCATCCCGTGGGAGAACCTGCCGGTGCAGGTGGTCGATCGCCTGACCCCCTGGCCT 26506

Qy       1236 -----GGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGAGCGGGACC 1290
Db      26507 CGGCGCGCAGAGGGCCCCCCGCGCCGTGCCGGCGTGTCGTCGTTCGGGATCAGCGGGACG 26566

Qy       1291 AACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCC--------GG 1342
Db      26567 AACGCGCATGTCATCCTCGAAGAAGCGCCGGCCGAGGCGCGGCGCGAGCCTGTCGAGGCG 26626
```

```
Qy       1343 CGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGGATG 1402
Db      26627 AGGCGGCGCCTGCGCTATTGCCGCTGGTGCTGTCGGGTCGAGACGAGGCCGCGGTGAATG 26686

Qy       1403 CACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCGATG 1462
Db      26687 CGCAGGCGGGGCGGTGGGCGAAGTGGCTGGAAGAGCACGGGGAGGTGGGGTGGTCGGACG 26746

Qy       1463 TGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGGCGA 1522
Db      26747 TGGTGCGCACGGCGGCGCTGCACCGGACGCACTTCGAGTCGCGGGCGTCGGTGCTTGCGG 26806

Qy       1523 CGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGCCCG 1582
Db      26807 CGAGCGCTGCGGGAGCTGTGGAGGGTCTTCGCGCGCTGTCGTCGGGGCGGCCGGATGCGG 26866

Qy       1583 GTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAGCTCGCCTTTCTCTTCACCGGAC 1642
Db      26867 CGGTGGTGAGCGGGACGGCGAA---GCGAGGCGGGAAGCTTGCGGTGCTGTTCACGGGGC 26923

Qy       1643 AGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGTTCCGCG 1702
Db      26924 AGGGCAGCCAGCGGCTCGGGATGGGGAAGAGGCTTTACGAAGTGTACCCCGTGTTCCGTG 26983

Qy       1703 AGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCCGCGAGG 1762
Db      26984 CGGCGTTCGACGAGGTGTGCGAGGCGCTGGACGCGCATCTCGACCGTGGGTTGAGAGAGG 27043

Qy       1763 TGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCTTCACCC 1822
Db      27044 TGGTGTTCGCGGCCGCGGGCAGCGAGGAAGGAGCGCAGCTGGAGCGGACGGAGTACACGC 27103

Qy       1823 AGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAG 1882
Db      27104 AGCCCGGGCTGTTTGCGCTGGAAGTGGCGCTGTACCGTCAGTGGGAGTCGTGGGGGCTGA 27163

Qy       1883 AGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCG 1942
Db      27164 AGCCCGCTGCGCTKCTGGGGCACTCGATAGGAGAGCTGAGCGCTGCGCACGTGGCGGGTG 27223

Qy       1943 TGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGC 2002
Db      27224 TGCTGAGCCTTGCGGACGCAGCGAAGCTAGTGTGCGCCCGCGGTCGGCTGATGCAGGGGT 27283

Qy       2003 TGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGG 2062
Db      27284 GCGAGGCCGGGGGAGCGATGGTGTCGGTGGAGGCCTCGGAGCCGGAGGTGCAGCGGGCGC 27343

Qy       2063 TGGCGCCGCACGCAGCGTCGG------TGTCGATCGCCGCGGTCAACGCTCCGGACCAGG 2116
Db      27344 TGTCGGAGGTCGGGGCGCAGGGGCGACTGAGCATCGCCGGGCTGAACGCGCCGATGCAGA 27403

Qy       2117 TGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCG 2176
Db      27404 CGGTGCTGAGCGGGGACGAAGCGGCGGTGCTCGCGGTGGCGCGACGGCTGGAGGCGCAGG 27463

Qy       2177 GGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGA 2236
Db      27464 GCCGGCGCACGCGGCGTCTGCGTGTGTCGCACGCGTTCCACAGCGCGCACATGGACGGGA 27523

Qy       2237 TGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCC 2296
Db      27524 TGCTGGAGGAGTTCGGGAAGGTGGCGCGGGAGTGCACGTACGCGCGGCCGCAGCTGGCGG 27583

Qy       2297 TGGTCAGCAATCTGA---GCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATT 2353
Db      27584 TGGTGAGCGGCGTGACGGGCGAGCTCGGTGGCGAAGAAGCGCTGATGTCGGCCGAGTACT 27643

Qy       2354 GGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGG 2413
Db      27644 GGGTGAGGCAGGTGCGCGAGGCGGTGCGCTTCCTGGACGGGATGCGCACGCTTGCGGCGG 27703

Qy       2414 CCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCTGGTGCCTG 2473
Db      27704 CGGGGGTGAGCACATACGTCGAGTGTGGGCCGGATGGCGTGCTGTGCGCGCTGGGGGCGG 27763

Qy       2474 CCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATC------GTCGCGCGCTGGGCGTG 2527
Db      27764 GGTGCCTGCCGGAGGGAGCCGAGGCGACGTTTGTGGCGAGCCTGCGGCGAGAGCAGGAGG 27823

Qy       2528 ACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCT 2587
Db      27824 AAGAGCGCGCGCTGGCGACGGCGGTGGCGACAGTGCACGTGCAGGGGCACGAGGTGGACT 27883

Qy       2588 CCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGC 2647
Db      27884 GGGCCCAGGTGCTGTCGGGCCGTGGCGGCCGGCCCGTGGAGCTGCCGACGTACGCGTTCC 27943

Qy       2648 AGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGACCGCCGTG 2707
Db      27944 AGCGGCAGCGCTACTGGCTGGAAGCGCCGAAGGCGCGTACCGACGTGGGCTCGGCGGGCT 28003

Qy       2708 CTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCGGC 2767
Db      28004 TGAGGGAGTCGGGGCATCCGCTGCTCGGAGCGGCAACGAAGCTGGCCGACGGCGACGGCC 28063
```

RESULT 10
AAF24892
ID   AAF24892 standard; DNA; 20394 BP.
XX
AC   AAF24892;
XX
DT   20-APR-2001 (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin; ss.
XX
OS   Streptomyces natalensis.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..20394
FT   /*tag= a
FT   /product= "polyketide synthase"
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999; 99EP-0201893.
XX
PA   (STAM) DSM NV.
XX
PI   Martin JF, Aparicio JF, Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   P-PSDB; AAB31558.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds -
XX
PS   Disclosure; Page 53-80; 116pp; English.
XX
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an

```
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

  Query Match             18.0%;  Score 766.6;  DB 22;  Length 20394;
  Best Local Similarity   53.6%;  Pred. No. 4.1e-114;
  Matches 1844;  Conservative    0;  Mismatches 1549;  Indels   48;  Gaps   10;

Qy         33 AGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCGATCTGAG 92
Db         99 AGAGCCCCTCGCGATCGTGTCGATGGCCTGCCGCTTCCCCGGCGGCGTCCGTTCGCCCGA 158

Qy         93 CGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCGCCGAAC- 151
Db        159 GGAGCTGTGGAGCTGCTGCGCGACGGCGTCGACGCGGTTTCCTCCTTCCCCCGTAACCG 218

Qy        152 --GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGACGCCCGT 209
Db        219 CGGCTGGGACCTGGACGCGCTCTACCACTCCGACCCGGCCCACCAGGGCACCAGCTATGC 278

Qy        210 TACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGCATCTC 269
Db        279 GCGCGAGGGCGGATTCCTGCATGACGCGGGCGAGTTCGACCCCGGCTTCTTCGGGATCTC 338

Qy        270 GCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGCTGGGA 329
Db        339 CCCGCGCGAGGCGCTCGCCATGGACCCCCAGCAGCGGCTGCTGCTGGAGACCGCATGGGA 398

Qy        330 GGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGAGTGTT 389
Db        399 AGCCGTCGAGCGGGCCGGTATCGACCCGGAGTCCCTCGCGGGCAGCCGAACGGGTGTCTT 458

Qy        390 CATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCCGCAGA 449
Db        459 TGTCGGCACCGGGCACGGAGGGTACGACGCCGAGGGCCGAC------GGCGTGCCGACGA 512

Qy        450 GATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATCTCGTA 509
Db        513 GGTCGGCGGGCACTTGCTGACGGGCAATCACATCAGCATCGCCTCCGGCCGGATTTCGTA 572

Qy        510 TGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCGCTGGT 569
Db        573 TGTCCTGGGGCTGGAAGGCCCTGCCCTGACCGTGGACACGGCCTGCTCCTCGTCGCTGGT 632

Qy        570 GGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTGGCTGG 629
Db        633 CGCCCTGCATCTGGCGATGCACGCGCTGCGGCGCGACGAATGCGCCATGGCCCTGGTGGG 692

Qy        630 TGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGGGCGCT 689
Db        693 CGGCGCGACCGTGATGTCCACGCCGCAGATGTTCGTGGAGTTCTCCCGCCAGCGCGGCCT 752

Qy        690 GGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGAGGCGA 749
Db        753 TGCCCCCGACGGCCGCTGCAAGCCGTTCGCGGCCGCCGCCGACGGCACCGGCTGGAGCGA 812

Qy        750 AGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGATCGGAT 809
Db        813 GGGTGTCGGACTGCTGCTCGTCGAGCGGCTCAGTGACGCCGTACGCAACGGCTATCCCGT 872

Qy        810 ATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTGACCGT 869
Db        873 CCTCGCCGTGCTGAAGGGCTCGGCCGTCAACCAGGACGGCGCGTCCAACGGCCTGACCGC 932

Qy        870 GCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGCTGCGC 929
Db        933 CCCCAACGGCCCCTCGCAGCAACGCGTCATCCGCCAGGCGCTGACCGGCGCGGGCCTCGC 992

Qy        930 CGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGACCCCAT 989
Db        993 CGCCTCGGACATCGACGCCGTGGAGGCGCACGGCACCGGCACCACCCTCGGCGACCCCGT 1052

Qy        990 CGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCGCTGCT 1049
Db       1053 CGAGGCGCACGCCCTGCTGGCCACCTACGGGCAGCAGCGCGCCGCCGACCGGCCGTGCGG 1112

Qy       1050 GATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACTGGGCT 1109
Db       1113 ACTCGGCTCCATGAAGTCCAACATCGGGCACACCCAGGCCGCCGCCGGTATCGCCGGCGT 1172

Qy       1110 GCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGCGCAGGC 1169
Db       1173 GATGAAAATGGTCCTGGCGATGCGGCACGGGCACCTGCCCAGGACCCTGCACCTGGACGA 1232

Qy       1170 GCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCGGACACC 1229
Db       1233 GCCCACCGGGCACGTCGACTGGAGCGAGGGCAACGCCAGGCTCCTCGCGGAGCCCGAGCC 1292

Qy       1230 GTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGAGCGGGAC 1289
Db       1293 CTGGCCGAGCGCCGGCCGGCCCCGTCGCGCCGCCGTCTCCTCCTTCGGCATCAGCGGCAC 1352

Qy       1290 CAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCG---------CC 1340
Db       1353 CAACGCCCACGTCATCCTGGAGCAGGCGCCCGCCCACGAGGCCGAACCGGCCCCCGAACC 1412

Qy       1341 GGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGGA 1400
Db       1413 GGCCGCCCGGCCGGGCGCGCTGCCCTGGATCCTGTCCGCCCGCACCGAAGCGGGCCTGCG 1472

Qy       1401 TGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCGA 1460
Db       1473 TGCCCAGGCCGACCGCCTCGGCCGCCACCTACGGGACCGCGCCGACCTCGAACCGGCCGC 1532

Qy       1461 TGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGGC 1520
Db       1533 CGTCGCCCATGCGCTCGCGGACACCCGGACCCTCATGGAACACCGCGCGGTAGTCGTCGC 1592

Qy       1521 GACGTCGAGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGCC 1580
Db       1593 GGGCGACCGGGAGGAGTTCCTGCGCGGCCTGGACGCCCTCGCCGCGGGCCGCACCGCCAA 1652

Qy       1581 CGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAGCTCGCCTTTCTCTTCACCGG 1640
Db       1653 CGGCCTGGTCAGCGGCGTCGCCGTCAAGGCCGCCAG------CGCGTTCCTCTTCGCCGG 1706

Qy       1641 ACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGTTCCG 1700
Db       1707 GCAGGGCTCCCAGCGACCGGGCATGGGGCGCGAACTGCACGCCGCGCACCCCGTGTTCGC 1766

Qy       1701 CGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCCGCGA 1760
Db       1767 CACGGCCTTCGACGCGGTGTGCGCCGAACTGGACCCACACCTGGACCGGCCGCTGCGCGA 1826

Qy       1761 GGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCTTCAC 1820
Db       1827 CATCGTCTTCGCCGAGGAAGGCAGCGCCGAGGCCGCCCTGCTCGACCAGACCGCCTACAC 1886

Qy       1821 CCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGGGTGT 1880
Db       1887 CCAGGCCGCGCTCTTCGCCCTGGAAACCGCCCTGTTCCGGCTCGTCGAATCCTGGGGCGT 1946

Qy       1881 AGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGGCGGG 1940
Db       1947 GGCACCCCGGTTCGTCGCCGGACACTCCATCGGCGAGCTGACCGCCGCCCACGTCAGTGG 2006

Qy       1941 CGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGCAGGC 2000
Db       2007 CGTGCTGACCCTCCACGACGCCGCACGGCTGGTCGCCGCGCGCGGCACCCTCATGCAGGC 2066

Qy       2001 GCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTGCTGC 2060
Db       2067 GCTGCCCGCAGGCGGCGCCATGGTGGCGGTCCAGGCCACCGAGGACGAGATCCGCGAGCG 2126

Qy       2061 GGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCCGGACCAGGTGGT 2120
```

```
Db       2127 TCTCGCCGGCCACGAGGACCACGTCGCCCTCGCGGCCGCCAACGGGCCCGATTCCACCGT 2186

Qy       2121 CATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCGGGGC 2180
Db       2187 CATTTCGGGCGACGAACAGGCCGTCACCGAGATCGCGGCCCACTGGGAGGCACAGGGCCG 2246

Qy       2181 GCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGATGCT 2240
Db       2247 CCGCACCAAGCGGCTGCGGGTCAGCCACGCCTTCCACTCCCCGCACATGGACGACATGCT 2306

Qy       2241 GGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCCTGGT 2300
Db       2307 GGAGGACTTCCGGCGCGTCGCCCGCGGTCTGACCTTCCACGCCCCCCGCATCCCCGTGGT 2366

Qy       2301 CAGCAATCTGA---GCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATTGGGT 2357
Db       2367 GTCCACGGTGACCGGCGCGCTCGCCACCGAAGACGAACTGCGCTCGCCCGACTACTGGGT 2426

Qy       2358 GCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGGCCGG 2417
Db       2427 GCGGCAGGTCCGCGAAACCGTCCGCTTCTGTGCTGCCGTGCGCACCCTTGAGGCCGAGGG 2486

Qy       2418 TGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCTGGTGCCTGCCTG 2477
Db       2487 CGTCACCACCTTCGTGGAGATCGGCACCGGCGGCGTCCTCACCCCCATGGTCCAGGACTG 2546

Qy       2478 CATG---CCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGACGAGCC 2534
Db       2547 TCTGACCACCCTCGAAGAGCCCGTTCTCGTCCCTCTGCTGCGCACCGGCCGCCCCGAAAC 2606

Qy       2535 GGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCCTGGGC 2594
Db       2607 CGTCGCTCTCACCGAGGGCGTCGCCACCGCCTTCGTGCACGGTGTCCCCGTCGACCGGTC 2666

Qy       2595 CGGCCTCTT---CCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGCAGCG 2651
Db       2667 CGCCTTCCCGGGCGCGCCCGGTACCTCCCGCGCGGACCTGCCCACCTACGCCTTCCAGCG 2726

Qy       2652 CGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGACCGCCGTGCTCC 2711
Db       2727 TCAGTGGTACTGGCTGGACCCGGCCGACCACGACGAGGGGGAGGCGGCCGCCGCCGAAGC 2786

Qy       2712 GGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCGGCGCAG 2771
Db       2787 GGGCGAGG-------CCGGATTCTGGGCGGCCGTCGAACGCGAGGACCTCCAGGAGCTGT 2839

Qy       2772 CGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTCGAGGCCGCCGG 2831
Db       2840 CGGCCGTCCTGGCCATCGACGGCAGCGAAGCGGACTCCCTCGGCAGCTTCCTGCCCACCC 2899

Qy       2832 CGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCACCTCGTGCTTCG 2891
Db       2900 TCTCCTCCTGGCGCAGGCAGCG---CAGGACCCAGGCCGCCGCGGACCGCTTCAGCTACC 2956

Qy       2892 GGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTCGACGCGGCGGG 2951
Db       2957 GCACCCACTGGGCCCCGCGCACCGCCT-----CGGGCGGCCCCACCGCCACCGGGCACTG 3011

Qy       2952 GCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGACCTGCCGGGAAA 3011
Db       3012 GCTCGTCGTCCTGCCCGAAGGCGGCACCGACGACCCGTGGACCGCCCGCCTCCTGGACGC 3071

Qy       3012 GCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTCGCCGTGGGCGA 3071
Db       3072 GCTGAACGACCAGGGCCTGCACACCGACGTACGCGAACTGCCCGCCGACCACGAGCCCGA 3131

Qy       3072 GGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCGGGAGCGTTTGC 3131
Db       3132 CGCCTGGGGCCGACACCCCGTGGACGGCGTGCTCTGTCTGCTGGCACTCGACGAGCGGCC 3191

Qy       3132 TACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCGCTCTCGGCGAT 3191

Db       3192 CACCCGCTCCTGCCCTCCGTACCGGCGCGGGCTGGCCGCCACCACCAACGCTGCTGCGCG 3251

Qy       3192 CGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTCGACAGAATAGC 3251
Db       3252 CCCTGAGGGCGCGGGCATCCAGGCACCGCTGTGGTGCGTGACCCGCGGCGCCGTCGCCGT 3311

Qy       3252 CCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGGGTCGGTCTCGC 3311
Db       3312 CGACCGGCACGAGGCGCTCAAGAGCCCCTTACAGGCACAGACATGGGGCCTGGGCCGGGT 3371

Qy       3312 CGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCCGGCACGCCCGA 3371
Db       3372 GGCGCCCCTGGAGTCCCCGCAGAGCTGGGGCGGGCTCATCGACCTGCCCGACAACCTGGA 3431

Qy       3372 GAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCCCGCTCGGACCG 3431
Db       3432 CGGACGGGCCGTCTCCGCGCTGCTGAGCACCCTCGCCGGGGAGGAGGACCAGGTCGCCGT 3491

Qy       3432 GTTCGTCGCCGACGTGCGCGC 3452
Db       3492 CCGCCCCGCCGGGGTCTTCGC 3512
```

RESULT 11
AAT80413

```
ID   AAT80413 standard; DNA; 43280 BP.
XX
AC   AAT80413;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             816..14243
FT                   /*tag=  a
FT                   /transl_except= (pos: 816..818, aa: Met)
FT                   /note= "ORF1 encodes protein shown in AAW22601"
FT   CDS             14351..19945
FT                   /*tag=  b
FT                   /transl_except= (pos: 14351..14353, aa: Met)
FT                   /note= "ORF2 encodes protein shown in AAW22602"
FT   CDS             20010..31199
FT                   /*tag=  c
FT                   /transl_except= (pos: 20010..20012, aa: Met)
FT                   /note= "ORF3 encodes protein shown in AAW22603"
FT   CDS             31232..36067
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW22604"
FT   CDS             36249..41774
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW22605"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22601, AAW22602, AAW22603, AAW22604, AAW22605.
XX
```

```
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 2; Pages 8-66; 220pp; English.
XX
CC   This sequence represents the tylactone synthase gene cluster of the
CC   invention. This sequence is also referred to as the tylG gene, and was
CC   isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC   proteins which direct the synthesis of the polyketide tylactone, isolated
CC   from Streptomyces fradiae. Tylactone is the basic building block of the
CC   antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC   type of carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

  Query Match             17.7%;  Score 755.2;  DB 18;  Length 43280;
  Best Local Similarity   56.0%;  Pred. No. 2.6e-112;
  Matches 1603;  Conservative    0;  Mismatches 1208;  Indels   53;  Gaps    7;

Qy         24 CGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGAT 83
Db      36350 CGCGGCGCGGGAGCCGATCGCCGTCGTCGGCATGGCCTGCCGCTACCCGGGCGGTGTCAC 36409

Qy         84 CGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCC 143
Db      36410 CGGCCCCGAGGAGCTGTGGGAGCTGGTGGCCGGAGGCCGGGACGCGATCGGGCCGTTCCC 36469

Qy        144 CGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAA 200
Db      36470 CGTGGACCGGGGCTGGGACGTGGCGTCGGTGTACGACCCGGATCCCGAGTCGAAGGGCAC 36529

Qy        201 GACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTT 260
Db      36530 CACGTACTGCCGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTT 36589

Qy        261 CGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGT 320
Db      36590 CGGCATCTCGCCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGT 36649

Qy        321 GTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAAC 380
Db      36650 GTCCTGGGAGGCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGG 36709

Qy        381 GGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGC 440
Db      36710 TGGTGTCTACGTGGGCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGA 36769

Qy        441 GTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCG 500
Db      36770 GGGCTCGGAGGGCTATCTGCTGACCGGCAGCGCCGACGCG------GTGATGTCCGGCCG 36823

Qy        501 AATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTC 560
Db      36824 CATCTCCTACGCGCTCGGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTC 36883

Qy        561 CTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGC 620
Db      36884 CTCGCTGGTGGCGCTGCATCTGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGC 36943

Qy        621 CCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGAC 680
Db      36944 GCTGGCGGGCGGGGTGGCGGTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCA 37003

Qy        681 CCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGG 740
Db      37004 GAAGGGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGG 37063

Qy        741 ACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGG 800
Db      37064 CTGGGCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGG 37123

Qy        801 CGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGG 860
Db      37124 GCACACGGTCCTCGGCCTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGG 37183

Qy        861 TCTGACCGTGCCGAACGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGC 920
Db      37184 GCTGACCGCGCCCAACGGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGC 37243

Qy        921 AGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGG 980
Db      37244 CGGGCTGTCCCCGGAGGACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGG 37303

Qy        981 TGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCAC 1040
Db      37304 CGACCCCATCGAGGCCGGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCA 37363

Qy       1041 GCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGAT 1100
Db      37364 CCCGCTGTGGCTCGGCTCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGT 37423

Qy       1101 CACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCA 1160
Db      37424 CGGCGGCGTCATCAAGATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCA 37483

Qy       1161 CGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGC 1220
Db      37484 CGCCGACGAGCCGACCCCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTC 37543

Qy       1221 CCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCAT 1280
Db      37544 CGAGGTGCCGTGGCAGCGGACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGT 37603

Qy       1281 GAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCC 1340
Db      37604 CGGCGGCACCAATGCCCATGTCGTCCTCGAAGAGGCACCCGCCCCGCCCGCGCCGGAACC 37663

Qy       1341 GGCGCCGGAGCGACC--------------------------------GGCAGAGCTGCT 1367
Db      37664 GGCCGGGGAGGCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGGCCCCTGGCCTG 37723

Qy       1368 GGTGCTGTCGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCA 1427
Db      37724 GGTGGTCTCCGGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCA 37783

Qy       1428 TCTGGAGACCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCG 1487
Db      37784 CCTCTCCCGCACCCCCGGGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCGCCACGCG 37843

Qy       1488 CAGCGCGATGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGC 1547
Db      37844 CGCAGCCTTTGACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGC 37903

Qy       1548 CCTGGACGCTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTC 1607
Db      37904 CCTGGACGCGTTGGCCGAAGGACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGA--- 37960

Qy       1608 CTCACGCGGCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGG 1667
Db      37961 CCGGGACGGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCCGGGATGGC 38020

Qy       1668 CCGTGGGCTGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCT 1727
Db      38021 CCACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTGACGGACCG 38080

Qy       1728 GTTCAACCAGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGT 1787
Db      38081 TCTCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGCTGGACGCCCGACCCGGCTCGCC 38140

Qy       1788 CGACGCCGCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATA 1847
Db      38141 CGAAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCGGCGCTCTTCGCCGTCGAGGT 38200

Qy       1848 TGCGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAG 1907
```

```
Db      38201 GGCGCTCCACCGGCTGCTGGAGCACTGGGGGATGCGCCCCGACCTGCTGCTGGGGCACTC 38260

Qy       1908 CATCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTT 1967
Db      38261 GGTGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTGCTCGATCTCGACGACGCCTGCGC 38320

Qy       1968 CCTGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTC 2027
Db      38321 GCTGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTGCCGCCCGGCGGCGCGATGGTCTC 38380

Qy       2028 GATCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTC 2087
Db      38381 CGTGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTGGCCGGCCGCGAGGACGCCGTCTG 38440

Qy       2088 GATCGCCGCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCA 2147
Db      38441 CGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGC 38500

Qy       2148 TGCGATCGCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCA 2207
Db      38501 CGAGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCA 38560

Qy       2208 TGCGTTCCACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTC 2267
Db      38561 CGCCTTCCACTCACCCCTGATGGACGGCATGCTCGCCGGATTCCGGGAGGTCGCCGCCGG 38620

Qy       2268 GGTGAGCTACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAG---GCTTG 2324
Db      38621 CCTGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCCACGGTCACGGGGCGGCCCGCCCG 38680

Qy       2325 CACAGACGAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTT 2384
Db      38681 CCCCGGTGAACTCACCGGCCCCGACTACTGGGTGGCCCAGGTCCGTGAGCCCGTGCGCTT 38740

Qy       2385 CGCGGATGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCC 2444
Db      38741 CGCGGACGCGGTCCGCACGGCACACCGCCTCGGAGCCCGCACCTTCCTGGAGACCGGCCC 38800

Qy       2445 GAAATCGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCT 2504
Db      38801 GGACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTGGAGGACGACACCGTGGCCCTGCT 38860

Qy       2505 CGCATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCT 2564
Db      38861 ---GCCGGCGATCCACAAGCCCGGCACCGCGCCGCACGGTCCGGCGGCTCCCGGCGCGCT 38917

Qy       2565 CTGGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGT 2624
Db      38918 GCGGGCGGCCGCCGCCGCGTACGGCCGGGGCGCCCGGGTGGACT--GGGCCGGGATGCAC 38975

Qy       2625 GCCGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGA 2684
Db      38976 GCCGACGGCCCCGAGGGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTTCCGGCAC 39035

Qy       2685 CGCGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGG 2744
Db      39036 CGCCGCTACTGGCTCGCCCCGGGCCGCGCGGCGGACACCGACGACTGGATGTACCGGATC 39095

Qy       2745 CGCGGTGCGCGGCGGCGACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGG 2804
Db      39096 GGCTGGGACCGGCTGCCGGCTGTGACCGGCGGGGCCCGGACCGCCGGCCGCTGGCTGGTG 39155

Qy       2805 ACGCCGGGAGAAGGTCGAGGCCGCCGGCGACCGTCCGTTCCGGC 2848
Db      39156 ATCCACCCCGACAGCCCGCGCTGCCGGGAGCTGTCCGGCCACGC 39199
```

RESULT 12
AAV21187
ID   AAV21187 standard; DNA; 53789 BP.
XX
AC   AAV21187;
XX
DT   24-JUL-1998 (first entry)
XX
DE   Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
XX
OS   Amycolatopsis mediterranei.
XX

```
FH   Key             Location/Qualifiers
FT   CDS             1825..15543
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= "polyketide synthase"
FT   CDS             15550..30759
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= "polyketide synthase"
FT   CDS             30895..36060
FT                   /*tag= c
FT                   /label= ORF_C
FT                   /product= "polyketide synthase"
FT   CDS             36259..41325
FT                   /*tag= d
FT                   /label= ORF_D
FT                   /product= "polyketide synthase"
FT   CDS             41373..51614
FT                   /*tag= e
FT                   /label= ORF_E
FT                   /product= "polyketide synthase"
FT   CDS             51713..5293
FT                   /*tag= f
FT                   /label= ORF_F
FT                   /product= "polyketide synthase"
```

XX
PN   WO9807868-A1.
XX
PD   26-FEB-1998.
XX
PF   18-AUG-1997; 97WO-EP04495.
XX
PR   20-AUG-1996; 96EP-0810551.
XX
PA   (NOVS) NOVARTIS AG.
XX
PI   Engel N, Schupp T, Toupet C;
XX
DR   WPI; 1998-169172/15.
DR   P-PSDB; AAW52845, AAW52846, AAW52847, AAW52848, AAW52849, AAW52850.
XX
PT   Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT   to produce rifamycin and rifamycin analogues
XX
PS   Claim 4; Page 53-102; 205pp; English.
XX
CC   The present sequence represents a Amycolatopsis mediterranei rifamycin
CC   synthesis gene cluster DNA fragment from the present invention. The
CC   DNA fragment comprises a DNA region involved directly or indirectly
CC   in the gene cluster responsible for rifamycin synthesis, including
CC   the adjacent DNA regions to the right and left which, by reason of
CC   their function in connection with rifamycin biosynthesis, qualify
CC   as constituents of this rifamycin gene cluster, and functional
CC   fragments, derivatives or constituents of these. The Amycolatopsis
CC   mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC   for producing rifamycin, rifamycin analogues or precursors. It can also
CC   be used for inactivating or modifying genes involved in ansamycin or
CC   rifamycin biosynthesis. The DNA can be used for constructing mutant
CC   actinomycetes strains from which the natural rifamycin or ansamycin
CC   biosynthesis gene cluster has been partly or completely deleted. The
CC   DNA fragment can be used for assembling a library of polyketide
CC   synthases, which can be used for assembling a library of polyketides.
CC   A hybridisation probe of the invention can be used for identifying DNA

```
CC   fragments involved in the biosynthesis of ansamycins.
XX
SQ   Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

  Query Match              17.5%;  Score 745.8;  DB 19;  Length 53789;
  Best Local Similarity    56.3%;  Pred. No. 8.3e-111;
  Matches 1674;  Conservative    0;  Mismatches 1232;  Indels   69;  Gaps   12;

Qy         28 GCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCGAT 87
Db      41457 GCCGACGAGCCGCTCGCCGTCGTCGGGATGGCCTGCCGGTTCCCGGGTGGGTGTCCTCG 41516

Qy         88 CTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCGCC 147
Db      41517 CCCGAAGACCTGTGGCAGCTCGTGGCCGGCGGGGTCGACGCCCTTTCGGACTTCCCCGAC 41576

Qy        148 GAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGACG 204
Db      41577 GACCGGGGCTGGGAGCTGGACGGCCTGTTCGACCCGGACCCCGACCACCCCGGGACGTCG 41636

Qy        205 CCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGC 264
Db      41637 TACACCAGCCAGGGCGGCTTCCTGCGTGGCGCCGGGCTGTTCGACGCGGGCCTGTTCGGC 41696

Qy        265 ATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGC 324
Db      41697 ATCTCGCCGCGCGAGGCCCTCGTCATGGACCCGCAGCAGCGGGTGCTGCTGGAGACGTCG 41756

Qy        325 TGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGA 384
Db      41757 TGGGAGGCCCTCGAAGACGCCGGGGTCGACCCGCTTTCGCTGAAGGGCAGCGACGTCGGC 41816

Qy        385 GTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCC 444
Db      41817 GTGTTCTCCGGCGTCTTCACCCAGGGCTACGGCGCCGGGG------------CGATCACG 41864

Qy        445 GCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATC 504
Db      41865 CCGGACCTCGAGGCGTTCGCGGGCATCGGGGCGGCGTCGAGCGTGGCGTCGGGCCGGGTG 41924

Qy        505 TCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCG 564
Db      41925 TCCTACGTCTTCGGGCTCGAAGGACCGGCGGTCACCATCGACACCGCGTGTTCGTCGTCG 41984

Qy        565 CTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTG 624
Db      41985 CTGGTGGCCATCCACCTCGCCGCGCAGGCCCTGCGCGCGGGCGAGTGCTCGATGGCGCTC 42044

Qy        625 GCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGG 684
Db      42045 GCCGGCGGGGCGACGGTGATGCCGACGCCCGGCACCTTCGTCGCGTTCTCGCGGCAGCGG 42104

Qy        685 GCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGA 744
Db      42105 GTGCTGGCTGCCGACGGCCGGTCCAAGGCCTTCTCCTCGACCGCGGACGGCACCGGCTGG 42164

Qy        745 GGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGAT 804
Db      42165 GCCGAGGGCGCCGGGGTGCTCGTCCTCGAACGGCTTTCGGTCGCGCAGGAGCGCGGCCAC 42224

Qy        805 CGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTG 864
Db      42225 CGGATTCTCGCCGTGCTGCGCGGCAGCGCGGTCAACCAGGATGGCGCCTCCAACGGCCTG 42284

Qy        865 ACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGC 924
Db      42285 ACCGCGCCGAACGGGCCTTCGCAGCAGCGGGTGATCCGCAAGGCGCTCGCGGGCGCCGGG 42344

Qy        925 TGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGAC 984
Db      42345 CTGGTCGCGTCCGATGTGGACGTCGTGGAGGCGCACGGCACGGGCACCGCGCTGGGCGAC 42404

Qy        985 CCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCG 1044
Db      42405 CCGATCGAAGCGCAGGCGCTGCTGGCGACCTACGGCCAGGGCCGTGA------GCGGCCG 42458

Qy       1045 CTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACT 1104
Db      42459 CTGTGGCTGGGGTCGGTCAAGTCGAACTTCGGGCACACGCAGGCGGCCGCCGGGGTCGCG 42518

Qy       1105 GGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGCG 1164
Db      42519 GGCGTGATCAAGATGGTCCAGGCCCTGCGGCACGGCGCCATGCCGCCGACCCTGCACGTG 42578

Qy       1165 CAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCGG 1224
Db      42579 GCCGAGCCGACGCCGGAGGTCGACTGGTCGGCCGGTGCGGTGGAACTGCTGACCGAGCCG 42638

Qy       1225 ACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGAGC 1284
Db      42639 CGCGAGTGGCCCGCCGGTGATCGGCCGCGCCGGGCCGGGGTGTCCGCGTTCGGGATCAGC 42698

Qy       1285 GGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCGGCG 1344
Db      42699 GGGACGAACGCCCACCTGATCCTGGAGGAGGCGCCCCCGGCCGACGCGGTCGCGGAAGAA 42758

Qy       1345 CCGGAGCGACCGG---CAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGGAT 1401
Db      42759 CCGGAGTTCAAGGGGCCGGTGCCGCTGGTCGTCTCGGCGGGCAGCCCCACATCTTTGGCG 42818

Qy       1402 GCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCGAT 1461
Db      42819 GCTCAGGCCGGCCGGCTCGCGGAGGTCCTGG---CGTCCGGTGGTGTGTCCCGGGCCCGG 42875

Qy       1462 GTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGGCG 1521
Db      42876 CTGGCGAGCGGGCTGCTGTCGGGCCGGGCGCTGCTCGGTGACCGCGCGGTCGTGGTCGCG 42935

Qy       1522 ACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGCCC 1581
Db      42936 GGAACGGACGAGGACGCGGTGGCCGGGTTGCGTGCGCTGGCCCGCGGGGACCGCGCGCCC 42995

Qy       1582 GGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAGCTCGCCTTTCTCTTCACCGGA 1641
Db      42996 GGCGTGCTGACCGGTTCGGCCAA------GCACGGCAAGGTCGTCTACGTCTTCCCCGGC 43049

Qy       1642 CAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGTTCCGC 1701
Db      43050 CAGGGTTCGCAGCGGCTCGGGATGGGCCGCGAGCTCTACGACCGGTACCCGGTGTTCGCG 43109

Qy       1702 GAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCCGCGAG 1761
Db      43110 ACGGCGTTCGACGAGGCTTGCGAGCAGCTGGACGTCTGTCTGGCCGGCCGTGCCGGGCAC 43169

Qy       1762 GTGATGTGGCCGAACCGGCCAG---------CGTCGACGCCGCGCTGCTCGACCAGACA 1812
Db      43170 CGCGTGCGGGACGTCGTGCTCGGCGAAGTGCCCGCCGAAACCGGGCTGCTGAACCAGACG 43229

Qy       1813 GCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCG 1872
Db      43230 GTCTTCACCCAAGCCGGGCTGTTCGCGGTGGAGAGCGCGCTGTTCCGGCTCGCCGAATCC 43289

Qy       1873 TGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGC 1932
Db      43290 TGGGGTGTCCGGCCGGACGTGGTGCTCGGCCACTCCATCGGGGAGATCACCGCCGCGTAT 43349

Qy       1933 GTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTG 1992
Db      43350 GCCGCGGGCGTCTTCTCGCTGCCGGACGCCGCCCGGATCGTCGCGGCGCGCGGCCGGCTG 43409

Qy       1993 ATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTG 2052
Db      43410 ATGCAGGCGCTGGCGCCGGGCGGGGCGATGGTCGCCGTCGCCGCCTCCGAAGCCGAGGTG 43469

Qy       2053 GCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCCGGAC 2112
```

```
Db      43470 GC---------CGAACTGCTCGGCGACGGCGTGGAACTCGCCGCCGTCAACGGCCCTTCG 43520

Qy       2113 CAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCG 2172
                ||| ||| |   | || | || | |  | ||     |||  ||| ||     |||  || |
Db      43521 GCGGTAGTCCTTTCCGGGGACGCGGACGCGGTCGTCGCGGCCGCCGCCCGCATGCGCGAG 43580

Qy       2173 CGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCC 2232
              ||||||      ||||||  |||| | || ||||| ||||||||||   |||   |||||
Db      43581 CGCGGGCACAAGACCAAGCAGCTCAAGGTTTCGCACGCGTTCCACTCCGCGCGGATGGCG 43640

Qy       2233 CCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATC 2292
              |||||||||| || |||||      |||| |  | |||   |  ||  ||||   ||||
Db      43641 CCGATGCTGGCGGAGTTCGCCGCCGAGCTGGCCGGCGTGACGTGGCGCGAGCCGGAGATC 43700

Qy       2293 GTCCTGGTCAGCAATCTGA---GCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGC 2349
                  ||||| |||  |||    || ||   ||       | | |||   ||| |  |||||||
Db      43701 CCGGTGGTCTCCAACGTGACCGGCCGGTTCGCCGAGCCCGGCGAACTGACCGAGCCGGGC 43760

Qy       2350 TATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCAC 2409
              || ||||       |:|| |||     |   |||||| ||||| || || ||     |||     |
Db      43761 TACTGGGCCGAGCACGTGCGGCGGCCGGTGCGGTTCGCCGAGGGCGTCGCGGCCGCGACG 43820

Qy       2410 GCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCTGGTG 2469
              | | |||| |              ||||| ||| |||| |||       || |||||  || ||| ||
Db      43821 GAGTCCGGCGGCTCGCTGTTCGTGGAGCTCGGGCCGGGGGCGGCGCTGACCGCCCTCGT- 43879

Qy       2470 CCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGAC 2529
                     |   |   |     | ||   |       |    ||||  |     |||  |   |   ||
Db      43880 --------CGAGGAGACGGCCGAGGTCACCTGCGTCGCGGCCCTGCGGGACGACCGCCCG 43931

Qy       2530 GAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCC 2589
              |||     |   |   || ||    ||| ||| || |||  |  |  |   ||| |     ||||
Db      43932 GAGGTCACCGCGCTGATCACCGCGGTCGCCGAGCTGTTCGTCCGCGGGGTTGCGGTCGAT 43991

Qy       2590 TGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGCAG 2649
              ||| |  |  |||   | ||   |  |          ||    ||      ||||| ||||| |  |     |
Db      43992 TGGCCGGCCCTGCTGCCGCCGGTCACCGGGTTCGTCGACCTGCCGAAGTACGCCTTCGAC 44051

Qy       2650 CGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGACCGCCGTGCT 2709
              |    ||| ||| ||| |   |   |      ||   |  |  | |||   || |||   |     || ||
Db      44052 CAGCAGCACTATTGGCTGCAGCCCGCCGCGCAGGCCACGG-ACGCGGCCTCGCTCGGGCA 44110

Qy       2710 CCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCGGCGC 2769
                    | |   | |      ||   | |     | || ||   |||        ||  | ||   ||||
Db      44111 GGTCGCGGCCGACCACCCGCTGCTGGGCGCGGTGGTCCGGCTGCCGCAGTCGGACGGCCT 44170

Qy       2770 AGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTCGAGGCCGCC 2829
                 |   ||  ||||     | ||||  | |  |  |  |   |  ||  |||  |||  |  |
Db      44171 GGTCTTCACCTCG-----CGGCTGTCATTGAAATCGCACCCGTGGCTGGCCGACCACGTC 44225

Qy       2830 GGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCACCTCGTGCTT 2889
                 || | |   | | |    ||| |   |    | ||   | || ||| |       ||  |
Db      44226 ATCGGCGGGGTGGTGCTCGTCGCGGGCACCGGGCTCGTCGAGCTGGCCGTCCGGGCCGGG 44285

Qy       2890 CGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTCGACGCGGCG 2949
                    |   | |   | | | | || |       ||   |||||    |  | | | ||   |
Db      44286 GACGAGGCCGGCTGCCCGGTCCTCGAAGAACTCGTCATCGAGGCTCCGCTGGTCGTCCCC 44345

Qy       2950 GGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGG 2984
              |. | | | ||             ||||| |||    ||||
Db      44346 GACCACGGCGGGGTCCGGATCCAGGTCGTCGTGGG 44380
```

RESULT 13
AAZ87283

```
ID   AAZ87283 standard; DNA; 15872 BP.
XX
AC   AAZ87283;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag= a
FT                   /product= "vep ORF 1 amino acid sequence #1 (AAY77177)"
FT   CDS             14056..14151
FT                   /*tag= b
FT                   /product= "vep ORF 1 amino acid sequence #3 (AAY77199)"
FT   CDS             14167..15827
FT                   /*tag= c
FT                   /product= "vep ORF 1 amino acid sequence #2 (AAY77178)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77177, AAY77178, AAY77199.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Example 3; Figure 23; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants. The
CC   present sequence represents a Streptomyces venezuelae ATCC 15439 DNA
CC   sequence, designated vep ORF 1 in the specification, which actually
CC   contains 3 open reading frames, which encode proteins AAY77177-Y77178 and
CC   AAY77199. The vep ORF 1 protein is defined in the specification as a PHA
CC   monomer synthase.
XX
SQ   Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;

  Query Match            17.4%;  Score 742;  DB 21;  Length 15872;
```

```
Best Local Similarity   55.7%;  Pred. No. 3.6e-110;
Matches 1594;  Conservative    0;  Mismatches 1215;  Indels   53;  Gaps    7;

Qy        26 CAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCG 85
Db      8490 CCGACGACGATCCGATCGCGATCGTGGGCATGGCATGCCGCTATCCGGGTGGTGTGTCGT 8549

Qy        86 ATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCG 145
Db      8550 CGCCGGAGGACCTGTGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCG 8609

Qy       146 CCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGA 202
Db      8610 TCAACCGCGGCTGGGACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCA 8669

Qy       203 CGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCG 262
Db      8670 CGTACTGCCGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCG 8729

Qy       263 GCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGT 322
Db      8730 GCATCTCGCCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGT 8789

Qy       323 GCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGG 382
Db      8790 CCTGGGAGGCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTG 8849

Qy       383 GAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGT 442
Db      8850 GTGTCTACGTGGGCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGG 8909

Qy       443 CCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAA 502
Db      8910 GCTCGGAG------GGCTATCTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCA 8963

Qy       503 TCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCT 562
Db      8964 TCTCCTACGCGCTCGGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCT 9023

Qy       563 CGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCC 622
Db      9024 CGCTGGTGGCGCTGCATCTGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGC 9083

Qy       623 TGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCC 682
Db      9084 TGGCGGGCGGGGTGGCGGTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGA 9143

Qy       683 GGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGAC 742
Db      9144 AGGGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCT 9203

Qy       743 GAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCG 802
Db      9204 GGGCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGC 9263

Qy       803 ATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTC 862
Db      9264 ACACGGTCCTCGGCCTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGC 9323

Qy       863 TGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAG 922
Db      9324 TGACCGCGCCCAACGGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCG 9383

Qy       923 GCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTG 982
Db      9384 GGCTGTCCCCGGAGGACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCG 9443

Qy       983 ACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGC 1042
Db      9444 ACCCCATCGAGGCCGGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACC 9503

Qy      1043 CGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCA 1102
Db      9504 CGCTGTGGCTCGGCTCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCG 9563

Qy      1103 CTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACG 1162
Db      9564 GCGGCGTCATCAAGATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACG 9623

Qy      1163 CGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCC 1222
Db      9624 CCGACGAGCCGACCCCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCG 9683

Qy      1223 GGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGA 1282
Db      9684 AGGTGCCGTGGCAGCGGACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCG 9743

Qy      1283 GCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCGG 1342
Db      9744 GCGGCACCAATGCCCATGTCGTCCTCGAAGAGGCACCCGCCCCGCCCGCGCCGGAACCGG 9803

Qy      1343 CGCCGGAGCGACC--------------------------------GGCAGAGCTGCTGG 1369
Db      9804 CCGGGGAGGCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGGCCCCTGGCCTGGG 9863

Qy      1370 TGCTGTCGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATC 1429
Db      9864 TGGTCTCCGGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCACC 9923

Qy      1430 TGGAGACCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCA 1489
Db      9924 TCTCCCGCACCCCCGGGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCGCCACGCGCG 9983

Qy      1490 GCGCGATGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCC 1549
Db      9984 CAGCCTTTGACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGCCC 10043

Qy      1550 TGGACGCTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCT 1609
Db     10044 TGGACGCGTTGGCCGAAGGACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGA---CC 10100

Qy      1610 CACGCGGCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCC 1669
Db     10101 GGGACGGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCCGGGATGGCCC 10160

Qy      1670 GTGGGCTGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGT 1729
Db     10161 ACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTGACGGACCGTC 10220

Qy      1730 TCAACCAGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCG 1789
Db     10221 TCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGCTGGACGCCCGACCCGGCTCGCCCG 10280

Qy      1790 ACGCCGCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATG 1849
Db     10281 AAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCGGCGCTCTTCGCCGTCGAGGTGG 10340

Qy      1850 CGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCA 1909
Db     10341 CGCTCCACCGGCTGCTGGAGCACTGGGGGATGCGCCCCGACCTGCTGCTGGGGCACTCGG 10400

Qy      1910 TCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCC 1969
Db     10401 TGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTGCTCGATCTCGACGACGCCTGCGCGC 10460

Qy      1970 TGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGA 2029
Db     10461 TGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTGCCGCCCGGCGGCGCGATGGTCTCCG 10520

Qy      2030 TCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGA 2089
Db     10521 TGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTGGCCGGCCGCGAGGACGCCGTCTGCG 10580

Qy      2090 TCGCCGCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATG 2149
Db     10581 TCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCG 10640
```

```
Qy       2150 CGATCGCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATG 2209
Db      10641 AGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCACG 10700

Qy       2210 CGTTCCACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGG 2269
Db      10701 CCTTCCACTCACCCCTGATGGACGGCATGCTCGCCGGATTCCGGGAGGTCGCCGCCGGCC 10760

Qy       2270 TGAGCTACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAG---GCTTGCA 2326
Db      10761 TGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCCACGGTCACGGGGCGGCCCGCCCCGCC 10820

Qy       2327 CAGACGAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCG 2386
Db      10821 CCGGTGAACTCACCGGCCCCGACTACTGGGTGGCCCAGGTCCGTGAGCCCGTGCGCTTCG 10880

Qy       2387 CGGATGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGA 2446
Db      10881 CGGACGCGGTCCGCACGGCACACCGCCTCGGAGCCCGCACCTTCCTGGAGACCGGCCCGG 10940

Qy       2447 AATCGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCG 2506
Db      10941 ACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTGGAGGACGACACCGTGGCCCTGCTGC 11000

Qy       2507 CATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCT 2566
Db      11001 ---CGGCGATCCACAAGCCCGGCACCGCGCCGCACGGTCCGGCGGCTCCCGGCGCGCTGC 11057

Qy       2567 GGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGC 2626
Db      11058 GGGCGGCCGCCGCCGCGTACGGCCGGGGCGCCCGGGTGGACT--GGGCCGGGATGCACGC 11115

Qy       2627 CGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACG 2686
Db      11116 CGACGGCCCCGAGGGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTTCCGGCACCG 11175

Qy       2687 CGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCG 2746
Db      11176 CCGCTACTGGCTCGCCCCGGGCCGCGCGGCGGACACCGACGACTGGATGTACCGGATCGG 11235

Qy       2747 CGGTGCGCGGCGGCGACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGAC 2806
Db      11236 CTGGGACCGGCTGCCGGCTGTGACCGGCGGGGCCCGGACCGCCGGCCGCTGGCTGGTGAT 11295

Qy       2807 GCCGGGAGAAGGTCGAGGCCGCCGGCGACCGTCCGTTCCGGC 2848
Db      11296 CCACCCCGACAGCCCGCGCTGCCGGGAGCTGTCCGGCCACGC 11337
```

```
RESULT 14
AAI99682/c
ID   AAI99682 standard; DNA; 4411529 BP.
XX
AC   AAI99682;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ -
XX
PS   Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

  Query Match             17.4%;  Score 740.8;  DB 22;  Length 4411529;
  Best Local Similarity   56.7%;  Pred. No. 4.1e-110;
  Matches 1571;  Conservative    0;  Mismatches 1127;  Indels   74;  Gaps    8;

Qy         19 GAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGC 78
Db    3297714 GAGCAACGCGCGACAGAACCGGTGGCGGTGGTGGGAATCGGCTGCCGATTTCCGGGCGGC 3297655

Qy         79 GTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGA 138
Db    3297654 GTGGATGGCCCGGACGGACTCTGGGATGTCGTTTCGGCGGGGCGCGACGTGGTGTCGGAG 3297595

Qy        139 GTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCG 195
Db    3297594 TTTCCGACCGATCGGGGCTGGGATGTTGAGGGGCTATACGATCCGGATCCCGACGCTGAG 3297535

Qy        196 GGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCC 255
Db    3297534 GGCAAAACCTACACGCGATGGGGGGCATTTCTAGACGACGCCACCGGCTTTGACGCCGGG 3297475

Qy        256 TTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTG 315
Db    3297474 TTCTTCGGCATCGCTCCTAGTGAAGTGTTGGCGATGGATCCTCAGCAGCGGTTGATGCTG 3297415

Qy        316 GAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACG 375
Db    3297414 GAGGTTTCGTGGGAGGCGTTGGAGCACGCGGGGATTGACCCGTTGTCGTTGCGTGGTTCG 3297355

Qy        376 GAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCG 435
Db    3297354 GCGACCGGGGTGTACACCGGCATTTTCGCGGCGAGTTACG--------------GCAAT 3297310

Qy        436 ACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCG 495
Db    3297309 AGGGATACCGGGGGGCTGCAAGGGTACGGGTTGACCGGCACGTCAATCAGCGTGGCCTCG 3297250

Qy        496 GGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTAT 555
Db    3297249 GGACGTGTGTCGTACGTATTGGGGCTGCAGGGCCCGGCGGTGTCAGTGGATACCGCTTGT 3297190

Qy        556 TCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCC 615
Db    3297189 TCGTCGTCGTTGGTGGCGATCCATTGGGCGATGTCGTCATTGCGCTCCGGGGAGTGTGAT 3297130
```

```
Qy        616 ACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCG 675
Db    3297129 TTGGCGTTGGCCGGTGGGGTGACGGTGATGGGGCTGCCGTCGATTTTTGTGGGGTTTTCC 3297070

Qy        676 AAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGG 735
Db    3297069 CGGCAGCGGGGGTTGGCTGCCGATGGGCGATGTAAGGCGTTTGCGGCCGCGGCCGACGGC 3297010

Qy        736 TTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCG 795
Db    3297009 ACCGGGTGGGGCGAGGGCGCCGGGGTGGTTGTGCTGGAGCGGCTATCGGACGCCCGGCGG 3296950

Qy        796 GACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGC 855
Db    3296949 TTGGGGCATTCGGTGCTGGCGGTGGTGCGGGGGAGCGCGGTCAACCAGGATGGTGCGTCC 3296890

Qy        856 AGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCG 915
Db    3296889 AATGGGCTGACCGCGCCCAATGGGCTCGCCCAGCAGCGAGTGATCCAGGTGGCGTTGGCC 3296830

Qy        916 GACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACG 975
Db    3296829 AACGCCGGGTTGTCCGCCGCAGACGTAGACGTGGTGGAGGCACACGGGACCGCCACCACG 3296770

Qy        976 CTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTC 1035
Db    3296769 TTGGGTGATCCCATCGAAGCGCAGGCATTGCTGTCCACCTACGGGCAGGGAGGTCCAGCG 3296710

Qy       1036 GCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCG 1095
Db    3296709 GAGCAGCCGCTGTGGGTGGGCTCAATCAAGTCGAACATGGGCCACACCCAGGCCGCCGCG 3296650

Qy       1096 GGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCAC 1155
Db    3296649 GGGGTCGCCGGGGTGATCAAGATGGTGCAGGCGATGCGTCATGGGGTGATGCCGGCGACG 3296590

Qy       1156 CTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACG 1215
Db    3296589 TTGCATGTGGATGAGCCTTCCCCGCGGGTGGATTGGACCAGCGGCGCGGTGTCGGTGTTG 3296530

Qy       1216 CGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTC 1275
Db    3296529 ACCGAGGCCAGGGAGTGGTCGGTTGACGGGCGTCCGCGCCGGGCGGCGGTGTCCTCCTTT 3296470

Qy       1276 GGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGAC------- 1328
Db    3296469 GGGATCAGCGGCACCAACGCGCATCTGATCTTGGAAGAGGCCCCCGTACCAGCCCCCGCA 3296410

Qy       1329 ------------------GTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGG 1369
Db    3296409 GAGGCGCCGGTTGAGGCGAGCGAAAGCACCGGGGGCCGCGGCCGTCGATGGTGCCGTGGG 3296350

Qy       1370 TGCTGTCGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATC 1429
Db    3296349 TGATTTCGGCGAGATCGGCTGAGGCGTTGACGGCGCAGGCGGGTCGACTTATGGCCCACG 3296290

Qy       1430 TGGAGACCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCA 1489
Db    3296289 TGCAGGCCAACCCAGGGCTGGATCCGATCGATGTGGGGTGCTCGTTGGCCAGT---CGCT 3296233

Qy       1490 GCGCGATGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCC 1549
Db    3296232 CGGTGTTTGAGCACCGAGCGGTGGTGGTCGGCGCAAGCCGTGAGCAACTGATTGCCGGGC 3296173

Qy       1550 TGGACGCTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCT 1609
Db    3296172 TGGCTGGGCTCGCGGCGGGCGAGCCGGGTGCCGGCGTGGCGGTCGGTCAGCCAG---GGT 3296116

Qy       1610 CACGCGGCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCC 1669
Db    3296115 CGGTGGGCAAGACGGTGGTCGTGTTTCCTGGGCAGGGCGCGCAGCGCATCGGGATGGGCC 3296056

Qy       1670 GTGGGCTGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGT 1729
Db    3296055 GCGAGTTGTACGGCGAGTTGCCCGTGTTTGCGCAGGCATTCGATGCGGTGGCCGACGAGT 3295996

Qy       1730 TCAACCAGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCGAACCGGCCAGCGTCG 1789
Db    3295995 TGGACCGGCATCTGCGGTTGCCGCTGCGCGACGTTATTTGGG-----------GTGCCG 3295948

Qy       1790 ACGCCGCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATG 1849
Db    3295947 ATGCGGATTTGCTTGACAGCACCGAATTTGCTCAGCCCGCGTTGTTCGCGGTGGAGGTGG 3295888

Qy       1850 CGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCA 1909
Db    3295887 CATCGTTCGCGGTGTTGCGGGATTGGGGTGTGCTTCCGGACTTCGTCATGGTCACTCCG 3295828

Qy       1910 TCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCC 1969
Db    3295827 TTGGAGAGCTGGCGGCGGCGCACGCGGCCGGTGTGTTGACGTTGGCGGACGCGGCGATGC 3295768

Qy       1970 TGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGA 2029
Db    3295767 TGGTGGTGGCGCGGGGCCGGTTGATGCAGGCGCTGCCGGCAGGCGGTGCGATGGTGGCGG 3295708

Qy       2030 TCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGA 2089
Db    3295707 TGGCTGCCAGTGAGGACGAGGT---------GGAGCCGCTGCTGGGTGAGGGTGTGGGGA 3295657

Qy       2090 TCGCCGCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATG 2149
Db    3295656 TCGCTGCGATCAACGCGCCCGAATCGGTGGTGATCTCCGGTGCGCAGGCCGCGGCAAATG 3295597

Qy       2150 CGATCGCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATG 2209
Db    3295596 CGATTGCGGATCGGTTCGCCGCGCAGGGTCGGCGGGTGCACCAGTTGGCGGTCTCGCATG 3295537

Qy       2210 CGTTCCACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGG 2269
Db    3295536 CGTTTCATTCGCCGTTGATGGAGCCGATGCTCGAGGAGTTCGCGCGTGTCGCGGCCCGGG 3295477

Qy       2270 TGAGCTACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAG 2329
Db    3295476 TGCAGGCACGCGAGCCCCAGCTTGGGCTGGTGTCGAACGTGACGGGCGAGTTGGCCGGCC 3295417

Qy       2330 ACGAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGG 2389
Db    3295416 CTGATTTCGGGTCGGCGCAGTACTGGGTGGACCACGTTCGTCGGCCGGTGCGCTTCGCGG 3295357

Qy       2390 ATGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAAT 2449
Db    3295356 ACAGTGCGCGTCATTTGCAGACCCTTGGGGCGACCCACTTCATCGAGGCCGGCCCGGGAA 3295297

Qy       2450 CGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCAT 2509
Db    3295296 GTGGTTTGACTGGCTCGATCGAGCAGTCCTTGGCCCCGGCTGAGGCGATGGTGGTGTCGA 3295237

Qy       2510 CGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGG 2569
Db    3295236 TGCTGGGCAAAGACCGGCCCGAGCTGGCCTCGGCGCTCGGTGCTGCCGGTCAGGTGTTCA 3295177

Qy       2570 CCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCC---CTCAGGGGGGCGGCGGGTGC 2626
Db    3295176 CCACCGGTGTGCCGGTGCAGTGGTCGGCGGTGTTCGCCGGCTCGGGTGGACGGCGGGTGC 3295117

Qy       2627 CGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACG 2686
Db    3295116 AGCTGCCCACGTATGCGTTTCAGCGACGGCGGTTTTGGGAGACGCCGGGCGCGGATGGGC 3295057

Qy       2687 CGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCG 2746
Db    3295056 CCGCCGATGCGGCCGGGTTGGGTCTGGGCGCGACCGAGCATGCCTTGTTGGGTGCGGTGG 3294997

Qy       2747 CGGTGCGCGGCG 2758
```

```
Db    3294996 TCGAGCGGCCCG 3294985

RESULT 15
AAT68715
ID   AAT68715 standard; DNA; 15872 BP.
XX
AC   AAT68715;
XX
DT   01-SEP-1997  (first entry)
XX
DE   Streptomyces venezuelae polyketide synthase vep ORF1.
XX
KW   Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW   polyhydroxybutyrate; biodegradable polymer; vep gene;
KW   metabolic engineering; ss.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag=  a
FT   CDS             14056..14136
FT                   /*tag=  b
FT   CDS             14148..15827
FT                   /*tag=  c
XX
PN   WO9722711-A1.
XX
PD   26-JUN-1997.
XX
PF   18-DEC-1996;   96WO-US20119.
XX
PR   19-DEC-1995;   95US-0008847.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Williams MD,  Xue Y;
XX
DR   WPI; 1997-341701/31.
DR   P-PSDB; AAW19629, AAW19630, AAW00918.
XX
PT   Expression cassettes for production of polyhydroxyalkanoate(s) -
PT   provide wide range of biodegradable polymers for medical or
PT   industrial use
XX
PS   Claim 54; Fig 23; 91pp; English.
XX
CC   Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
CC   synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
CC   also AAW19629-30 and AAW00918).  It contains 5 PKS modules, with a 5'
CC   loading module and a 3' end domain.  Each of the sequenced modules
CC   includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC   reductase and an acyl carrier protein domain.  The gene cluster was
CC   cloned using a heterologous hybridisation strategy from a genomic
CC   DNA library.  A novel expression cassette encoding the first module
CC   from the vep gene cluster and module 7 from the Streptomyces tylP
CC   gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC   activity and can be used for PHA prodn. in host (esp. insect) cells
CC   for use as a biodegradable polymer.
XX
SQ   Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

  Query Match             17.2%;  Score 735.6;  DB 18;  Length 15872;
  Best Local Similarity   55.6%;  Pred. No. 3.8e-109;
  Matches 1590;  Conservative    0;  Mismatches 1219;  Indels   53;  Gaps    7;

Qy         26 CAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATCG 85
Db       8490 CCGACGACGATCCAATCGCGATCGTGGGCATGGCATGCCGCTATCCAGGTGGTGTGTCGT 8549

Qy         86 ATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCCG 145
Db       8550 CGCCGGAGGACCTGTGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCG 8609

Qy        146 CCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGA 202
Db       8610 TCAACCGCGGCTGGGACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCA 8669

Qy        203 CGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCG 262
Db       8670 CGTACTGCCGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCG 8729

Qy        263 GCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGT 322
Db       8730 GCATCTCGCCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGGGGCTGCTGCTGGAGGTGT 8789

Qy        323 GCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGG 382
Db       8790 CCTGGGAGGCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTG 8849

Qy        383 GAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGT 442
Db       8850 GTGTCTACGTGGGCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCT------GGTGC 8903

Qy        443 CCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAA 502
Db       8904 CCGAGGGGTCGGAGGGCTATCTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCA 8963

Qy        503 TCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCT 562
Db       8964 TCTCCTACGCGCTCGGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCT 9023

Qy        563 CGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCC 622
Db       9024 CGCTGGTGGCGCTGCATCTGGCGGTACGGGCGCTGCGGCACGGCGAGGGCGGGCTCGCGC 9083

Qy        623 TGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCC 682
Db       9084 TGGCGGGCGGGGTGGCGGTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGA 9143

Qy        683 GGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGAC 742
Db       9144 AGGGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCT 9203

Qy        743 GAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCG 802
Db       9204 GGGCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGC 9263

Qy        803 ATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTC 862
Db       9264 ACACGGTCCTCGGCCTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGC 9323

Qy        863 TGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAG 922
Db       9324 TGACCGCGCCCAACGGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCG 9383

Qy        923 GCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTG 982
Db       9384 GGCTGTCCCCGGAGGACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCG 9443

Qy        983 ACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGC 1042
Db       9444 ACCCCATCGAGGCCGGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACC 9503

Qy       1043 CGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCA 1102
Db       9504 CGCTGTGGCTCGGCTCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCG 9563

Qy       1103 CTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACG 1162
Db       9564 GCGGCGTCATCAAGATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACG 9623

Qy       1163 CGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCC 1222
Db       9624 CCGACGAGCCGACCCCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCG 9683
```

51

```
Qy       1223 GGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGA 1282
Db       9684 AGGTGCCGTGGCAGCGGACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCG 9743

Qy       1283 GCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCGG 1342
Db       9744 GCGGCACCAATGCCCATGTCGTCCTCGAAGAGGCACCCGCCCCGCCCGCGCCGGAACCGG 9803

Qy       1343 CGCCGGAGCGACC-------------------------------GGCAGAGCTGCTGG 1369
Db       9804 CCGGGGAGGCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGGCCCCTGGCCTGGG 9863

Qy       1370 TGCTGTCGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATC 1429
Db       9864 TGGTCTCCGGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCACC 9923

Qy       1430 TGGAGACCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCA 1489
Db       9924 TCTCCCGCACCCCCGGGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCGCCACGCGCG 9983

Qy       1490 GCGCGATGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCC 1549
Db       9984 CAGCCTTTGACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGCCC 10043

Qy       1550 TGGACGCTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCT 1609
Db      10044 TGGACGCGTTGGCCGAAGGACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGA---CC 10100

Qy       1610 CACGCGGCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCC 1669
Db      10101 GGGACGGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCCGGGATGGCCC 10160

Qy       1670 GTGGGCTGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGT 1729
Db      10161 ACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTGACGGACCGTC 10220

Qy       1730 TCAACCAGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCG 1789
Db      10221 TCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGCTGGACGCCCGACCCGGCTCGCCCG 10280

Qy       1790 ACGCCGCGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATG 1849
Db      10281 AAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCGGCGCTCTTCGCCGTCGAGGTGG 10340

Qy       1850 CGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCA 1909
Db      10341 CGCTCCACCGGCTGCTGGAGCACTGGGGATGCGCCCCGACCTGCTGCTGGGGCACTCGG 10400

Qy       1910 TCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCC 1969
Db      10401 TGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTGCTCGATCTCCACGACGCCTGCGCGC 10460

Qy       1970 TGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGA 2029
Db      10461 TGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTGCCGCCCGGCGGCGCGATGGTCTCCG 10520

Qy       2030 TCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGA 2089
Db      10521 TGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTGGCCGGCCGCGAGGACGCCGTCTGCG 10580

Qy       2090 TCGCCGCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATG 2149
Db      10581 TCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCG 10640

Qy       2150 CGATCGCGGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATG 2209
Db      10641 AGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCACG 10700

Qy       2210 CGTTCCACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGG 2269
Db      10701 CCTTCCACTCACCCCTGATGGACGGCATGCTCGCCGGATTCCGGGAGGTCGCCGCCGGCC 10760

Qy       2270 TGAGCTACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAG---GCTTGCA 2326
Db      10761 TGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCCACGGTCACGGGGCGGCCCGCCCGCC 10820

Qy       2327 CAGACGAGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCG 2386
Db      10821 CCGGTGAACTCACCGGCCCCGACTACTGGGTGGCCCAGGTCCGTGAGCCCGTGCGCTTCG 10880

Qy       2387 CGGATGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGA 2446
Db      10881 CGGACGCGGTCCGCACGGCACACCGCCTCGGAGCCCGCACCTTCCTGGAGACCGGCCCGG 10940

Qy       2447 AATCGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCG 2506
Db      10941 ACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTGGAGGACGACACCGTGGCCCTGCTGC 11000

Qy       2507 CATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCT 2566
Db      11001 ---CGGCGATCCACAAGCCCGGCACCGCGCCGCACGGTCCGGCGGCTCCCGGCGCGCTGC 11057

Qy       2567 GGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCTCAGGGGGGCGGCGGGTGC 2626
Db      11058 GGGCGGCCGCCGCCGCGTACGGCCGGGGCGCCCGGGTGGACT--GGGCCGGGATGCACGC 11115

Qy       2627 CGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACG 2686
Db      11116 CGACGGCCCCGAGGGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTTCCGGCACCG 11175

Qy       2687 CGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCG 2746
Db      11176 CCGCTACTGGCTCGCCCCGGGCCGCGCGGCGGACACCGACGACTGGATGTACCGGATCGG 11235

Qy       2747 CGGTGCGCGGCGGCGACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGAC 2806
Db      11236 CTGGGACCGGCTGCCGGCTGTGACCGGCGGGGCCCGGACCGCCGGCCGCTGGCTGGTGAT 11295

Qy       2807 GCCGGGAGAAGGTCGAGGCCGCCGGCGACCGTCCGTTCCGGC 2848
Db      11296 CCACCCCGACAGCCCGCGCTGCCGGGAGCTGTCCGGCCACGC 11337
```

Search completed: October 3, 2003, 12:27:59
Job time : 1144.88 secs

```
                  GenCore version 5.1.6
          Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         October  3, 2003, 11:40:41 ; Search time 15260.8 Seconds
                                        (without alignments)
                                        11435.853 Million cell updates/sec

Title:          US-10-014-717-1_COPY_7610_11875
Perfect score:  4266
Sequence:       1 gtggcggatcgtcccatcga..........gggaaatcattgccctatga 4266

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters:      5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4266	100.0	68750	1	AF210843	AF210843 Sorangium
	2	4266	100.0	68750	6	AR193029	AR193029 Sequence
	3	4266	100.0	68750	6	AR199551	AR199551 Sequence
	4	4266	100.0	68750	6	AR199559	AR199559 Sequence
	5	4266	100.0	68750	6	AR199567	AR199567 Sequence
	6	4266	100.0	68750	6	AR201097	AR201097 Sequence
	7	4266	100.0	68750	6	AR208671	AR208671 Sequence
	8	4187.6	98.2	58733	1	AF217189	AF217189 Sorangium
	9	4187.6	98.2	71989	6	AR172664	AR172664 Sequence
	10	1287.4	30.2	16124	6	AX024277	AX024277 Sequence
	11	928.8	21.8	10910	6	AX024276	AX024276 Sequence
	12	859	20.1	48000	1	POL505006	AJ505006 Sorangium
c	13	844.4	19.8	49736	1	AF319998	AF319998 Stigmatel
	14	817.2	19.2	66808	1	SAU421825	AJ421825 Stigmatel
	15	814.2	19.1	3978	6	AR266876	AR266876 Sequence
	16	814.2	19.1	12249	6	AR266931	AR266931 Sequence
	17	814.2	19.1	18331	6	AR266948	AR266948 Sequence
	18	812	19.0	42603	1	AF188287	AF188287 Stigmatel
	19	805	18.9	27541	6	AX211706	AX211706 Sequence
	20	805	18.9	123580	1	AF263912	AF263912 Streptomy
	21	805	18.9	125401	6	AX211739	AX211739 Sequence
	22	803.2	18.8	33529	6	AR166425	AR166425 Sequence
c	23	790.6	18.5	290850	1	SCO939127	AL939127 Streptomy
c	24	784	18.4	84985	1	SNA278573	AJ278573 Streptomy
	25	772	18.1	113193	1	AF357202	AF357202 Streptomy
	26	772	18.1	113193	6	AX703543	AX703543 Sequence
	27	770.8	18.1	5355	6	AX697995	AX697995 Sequence
	28	770.8	18.1	60196	6	AX697977	AX697977 Sequence
	29	766.6	18.0	20394	1	SNA132222	AJ132222 Streptomy
	30	766.6	18.0	20394	6	AX067996	AX067996 Sequence
c	31	766.4	18.0	39314	1	SGR300302	AJ300302 Streptomy
	32	765.8	18.0	69644	1	AY179507	AY179507 Streptomy
	33	761.2	17.8	104326	1	AB070940	AB070940 Streptomy
c	34	761.2	17.8	281450	1	AP005032	AP005032 Streptomy
	35	755.2	17.7	43280	1	SFU78289	U78289 Streptomyce
	36	745.8	17.5	53784	1	AMM223012	AJ223012 Amycolato
	37	745.8	17.5	53789	6	A69720	A69720 Sequence 3
	38	745.8	17.5	90445	1	AF040570	AF040570 Amycolato
c	39	741.4	17.4	318050	1	BX248344	BX248344 Mycobacte
c	40	740.8	17.4	16096	1	AE007124	AE007124 Mycobacte
c	41	740.8	17.4	24292	1	MTCY24G1	Z83858 Mycobacteri
	42	732.8	17.2	78210	1	AB070949	AB070949 Streptomy
c	43	732.8	17.2	300425	1	AP005022	AP005022 Streptomy
	44	726.2	17.0	65140	6	AX211705	AX211705 Sequence
	45	725	17.0	49377	6	I88042	I88042 Sequence 1

ALIGNMENTS

```
RESULT 1
AF210843
LOCUS       AF210843                68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
```

```
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="Polyangium cellulosum"
                     /mol_type="genomic DNA"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1..1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900..3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
                     /db_xref="GI:6724239"
                     /translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
                     AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
                     EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
                     SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
                     VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
                     GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
                     TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
                     IGIVMTQSAGFLFSGALERFWRSVYVATESA"
     CDS             3415..5556
                     /note="ORF3; similar to Na/H and drug/H antiporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26906.1"
                     /db_xref="GI:6724240"
                     /translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
                     GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLMAGIEVDVGILRKEA
                     RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
                     RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
                     LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
                     NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
                     LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
                     SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
                     IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
                     IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
                     ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
                     ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
                     YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
     CDS             complement(5612..5992)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26907.1"
                     /db_xref="GI:6724241"
                     /translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTIAYTDRE"
     CDS             6226..6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MKHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYANFELLITAEKQSSPQSPSSAAVIGPTSVG"
     gene            7610..11875
                     /gene="epoA"
     CDS             7610..11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
     misc_feature    7640..8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
     misc_feature    9236..10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
     misc_feature    10529..11428
                     /gene="epoA"
                     /note="Region: enoyl reductase"
     misc_feature    11549..11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
     gene            11872..16104
                     /gene="epoP"
     CDS             11872..16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
```

```
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRS"
     gene            16251..21749
                     /gene="epoB"
     CDS             16251..21749
                     /gene="epoB"
                     /note="EPOS B"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26920.1"

  Query Match            100.0%;  Score 4266;  DB 1;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109
```

```
Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy       1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy       1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy       1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy       2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy       2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy       2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy       2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy       2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy       2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy      3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy      3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy      3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 11509

Qy      3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy      3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy      4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy      4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy      4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy      4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy      4261 CTATGA 4266
             ||||||
Db     11870 CTATGA 11875

RESULT 2
AR193029
LOCUS       AR193029              68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 4266;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy        61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy       121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy       181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy       241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy       301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy       361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy       421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy       481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy       541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy       601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy       661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy       721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy       781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy       841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy       901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy       961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy      1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy      1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy      1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGTGATCTT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy       1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy       1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy       1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy       2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy       2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy       2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy       2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy       2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy       2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969
```

```
Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875
```

```
RESULT 3
AR199551
LOCUS       AR199551                68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 4266;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449
```

```
Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy       1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy       1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy       1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy       2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy       2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy       2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy       2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy       2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy       2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875
```

```
RESULT 4
AR199559
LOCUS       AR199559              68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 4266;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy       1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy       1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy       1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy       2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy       2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy       2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy       2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy       2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy       2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309
```

```
Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875
```

```
RESULT 5
AR199567
LOCUS       AR199567                68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 4266;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789
```

```
Qy    181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy    241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy    301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy    361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy    421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy    481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy    541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy    601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy    661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy    721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy    781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy    841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy    901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy    961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy   1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy   1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy   1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy   1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy   1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy   1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy   1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy   1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy   1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy   1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy   1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy   1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy   1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy   1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy   1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy   1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy   1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy   2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy   2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy   2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy   2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy   2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy   2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875
```

```
RESULT 6
AR201097
LOCUS       AR201097                 68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
```

```
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a   22456 c   25539 g   11159 t
ORIGIN

  Query Match             100.0%;  Score 4266;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy       1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy       1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy       1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649
```

```
Qy       2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy       2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy       2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy       2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy       2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy       2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy       2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy       2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy       2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy       2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy       3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGCGGCGAGGGAGTAGACGTC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGCGGCGAGGGAGTAGACGTC 11089

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11090 GTGCTCAAGTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy       3721 ACCCCTCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11330 ACCCCTCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 11509

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875


RESULT 7
AR208671
LOCUS       AR208671                 68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1..68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 4266;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 7669

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 7729

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 7789

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 7849

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 7909

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 7969

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 8029

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 8089

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 8209

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 8269

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 8329

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 8389

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 8449

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8450 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 8509

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8510 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 8569

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8570 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 8629

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8630 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 8689

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8690 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 8749

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8750 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 8809

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 8869

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 8929

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 8989

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 9049

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 9109

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 9169

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 9229

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 9289

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 9349

Qy     1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9350 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 9409

Qy     1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9410 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 9469

Qy     1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9470 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 9529

Qy     1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9530 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 9589

Qy     1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9590 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 9649

Qy     2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9650 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 9709

Qy     2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9710 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 9769

Qy     2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9770 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 9829

Qy     2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9830 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 9889

Qy     2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9890 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 9949

Qy     2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     9950 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 10009

Qy     2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10010 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 10069

Qy     2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10070 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 10129

Qy     2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10130 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 10189

Qy     2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10190 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 10249

Qy     2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10250 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 10309

Qy     2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10310 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 10369

Qy     2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10370 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 10429

Qy     2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10430 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 10489

Qy     2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10490 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 10549

Qy     2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10550 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 10609

Qy     3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10610 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 10669

Qy     3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10670 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 10729

Qy     3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10730 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 10789

Qy     3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10790 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 10849

Qy     3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10850 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 10909

Qy     3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10910 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 10969

Qy     3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10970 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 11029

Qy     3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11030 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 11089

Qy     3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11090 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 11149

Qy     3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11150 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 11209

Qy     3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11210 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 11269

Qy     3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11270 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 11329

Qy     3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11330 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 11389

Qy     3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11390 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 11449

Qy     3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11450 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 11509
```

```
Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11510 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 11569

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11570 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 11629

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11630 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 11689

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11690 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 11749

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11750 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 11809

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11810 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 11869

Qy       4261 CTATGA 4266
              ||||||
Db      11870 CTATGA 11875
```

```
RESULT 8
AF217189
LOCUS       AF217189               58733 bp    DNA     linear   BCT 09-JUN-2000
DEFINITION  Sorangium cellulosum putative transposase gene, partial cds;
            putative transposase gene, complete cds; epothilone biosynthesis
            gene cluster, complete sequence; putative membrane protein gene,
            complete cds.
ACCESSION   AF217189
VERSION     AF217189.1  GI:7453554
KEYWORDS    .
SOURCE      Polyangium cellulosum
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 58733)
  AUTHORS   Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and
            Julien,B.
  TITLE     Cloning and heterologous expression of the epothilone gene cluster
  JOURNAL   Science 287 (5453), 640-642 (2000)
  MEDLINE   20115953
   PUBMED   10649995
REFERENCE   2  (bases 1 to 58733)
  AUTHORS   Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
  TITLE     Isolation and characterization of the epothilone biosynthetic gene
            cluster from Sorangium cellulosum
  JOURNAL   Gene 249 (1-2), 153-160 (2000)
  MEDLINE   20293058
   PUBMED   10831849
REFERENCE   3  (bases 1 to 58733)
  AUTHORS   Julien,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
     source          1..58733
                     /organism="Polyangium cellulosum"
                     /mol_type="genomic DNA"
                     /strain="SMP44"
                     /db_xref="taxon:56"
     CDS             <1..992
                     /note="orfA"
                     /codon_start=3
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62878.1"
                     /db_xref="GI:7453555"
                     /translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
                     SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
                     DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
                     DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
                     GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
                     HIGAVRQVIDRRRSERHLPPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"
     CDS             989..1501
                     /note="orfB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62879.1"
                     /db_xref="GI:7453556"
                     /translation="MTDLTPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
                     RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
                     AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
                     RTKQRRTKKH"
     gene            1998..6263
                     /gene="epoA"
     CDS             1998..6263
                     /gene="epoA"
                     /note="type I polyketide synthase; epothilone PKS loading
                     module"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoA"
                     /protein_id="AAF62880.1"
                     /db_xref="GI:7453557"
                     /translation="MADRPIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
                     GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
                     PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
                     DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
                     QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"
     gene            6260..10492
                     /gene="epoB"
     CDS             6260..10492
                     /gene="epoB"
                     /note="non ribosomal peptide synthetase; epothilone PKS
                     module 1"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoB"
                     /protein_id="AAF62881.1"
                     /db_xref="GI:7453558"
                     /translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
                     PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
```

```
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
                     ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
                     STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
                     FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"
gene                 10639. .16137
                     /gene="epoC"
CDS                  10639. .16137
                     /gene="epoC"
                     /note="type I polyketide synthase; epothilone PKS module
                     2"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoC"
                     /protein_id="AAF62882.1"
                     /db_xref="GI:7453559"
                     /translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
                     ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
                     ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
                     HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
                     AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
                     GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
                     DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
                     ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
                     FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
                     REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
                     SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
                     VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
                     GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
                     EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
                     GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
                     VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
gene                 16134. .37907
                     /gene="epoD"
CDS                  16134. .37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
```

```
           GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
           FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
           RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
           LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
           LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
           DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
           AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
           NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
           AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
           GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
           AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
           RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA

  Query Match             98.2%;  Score 4187.6;  DB 1;  Length 58733;
  Best Local Similarity   98.9%;  Pred. No. 0;
  Matches 4217;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| | 
Db       1998 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGGGC 2057

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 2117

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db       2118 CGCGACACCGTCGGGCAAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 2177

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 GACCTCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 2237

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2238 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 2297

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2298 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 2357

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2358 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 2417

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2418 GCGCTGCCGCGAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 2477

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       2478 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGTCCTCGGGCTGCGAGGGCCGTGTGTCGCG 2537

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       2538 GTGGATACGGCCTATTCGTCCTCGCTCGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 2597

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2598 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 2657

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||| |||||||||| |||||||||||||||||| ||||||
Db       2658 CTCGTGTGGCTCTCGAAGACCCGCGCGCTGGCCACGGACGGTCGCTGCAAGGCGTTTTCG 2717

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2718 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 2777

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2778 AGTGGAGCCCGCGCGGACGGCGACCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 2837
```

```
Qy      841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
            |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2838 CACGACGGAGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 2897

Qy      901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2898 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 2957

Qy      961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2958 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 3017

Qy     1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
            ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     3018 CTCGGGCGAGACGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 3077

Qy     1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3078 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 3137

Qy     1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3138 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 3197

Qy     1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3198 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 3257

Qy     1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3258 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 3317

Qy     1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
            |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db     3318 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGGCCGGCAGAGCTGCTGGTGCTGTCGGCA 3377

Qy     1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
            ||||||||| ||||| |||||||||| |||||||||||||||||||||||||||||||||
Db     3378 AGGACCGCGGCAGCCTTGGATGCACACGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 3437

Qy     1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3438 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 3497

Qy     1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
            ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     3498 CACCGGCTCGCGGTGGCGGCGACGTCGAGCGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 3557

Qy     1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
            ||||||||||||||| ||||||||| ||||||| ||||||||||||||||||||||||||
Db     3558 GCGCAGGGACAGACGCCGCCCGGTGTGGTGCGCGGTATCGCCGATTCCTCACGCGGCAAG 3617

Qy     1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db     3618 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAT 3677

Qy     1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
            ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3678 GATGTATGGCCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 3737

Qy     1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3738 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 3797

Qy     1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
            ||||||||||||||||| ||||||||||||||||||||||||||| ||||||||||||||
Db     3798 CTCGACCAGACAGCCTTTACCCAGCCGGCGCTGTTCACCTTCGAGTATGCGCTCGCCGCG 3857

Qy     1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
            ||||||||||||||||| |||||||||||||||||||| |||||||||||||||||||||
Db     3858 CTGTGGCGGTCGTGGGGCGTAGAGCCGGAGTTGGTCGCTGGCCATAGCATCGGTGAGCTG 3917

Qy     1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3918 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 3977

Qy     1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
            |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     3978 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGCGGCGCCG 4037

Qy     2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4038 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 4097

Qy     2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
            |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4098 AACGGTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 4157

Qy     2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4158 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 4217

Qy     2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4218 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 4277

Qy     2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
            |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db     4278 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTGGCACAGACGAGGTGAGC 4337

Qy     2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4338 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 4397

Qy     2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4398 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 4457

Qy     2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
            |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db     4458 GGCCTGGTGCCTGCCTGCCTGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 4517

Qy     2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGCCGTCGGTGGC 2580
            |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db     4518 GGGCGTGACGAGCCAGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGCCGTCGGTGGC 4577

Qy     2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4578 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 4637

Qy     2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4638 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 4697

Qy     2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
            ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     4698 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGAAGGGGGGCGCGGTGCGCGGCGGC 4757

Qy     2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4758 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 4817

Qy     2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||| |
Db     4818 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTCGATCGC 4877

Qy     2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
            || |||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     4878 CTGGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTTGGCGAGGTCGAGATCGCCGTC 4937

Qy     2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4938 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 4997

Qy     3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4998 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 5057

Qy       3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
Db       5058 GCCGTGGGCGAGGGCGTGAACGGCCTTGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 5117

Qy       3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
Db       5118 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 5177

Qy       3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
Db       5178 CTCTCGGCGACCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 5237

Qy       3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
Db       5238 GACGGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCACGCGGCGACCGGCGGG 5297

Qy       3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
Db       5298 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 5357

Qy       3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
Db       5358 GGCACGCCCGAGAAGCGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 5417

Qy       3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
Db       5418 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 5477

Qy       3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
Db       5478 GTGCTCAACTCGCTTTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 5537

Qy       3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
Db       5538 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 5597

Qy       3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
Db       5598 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 5657

Qy       3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
Db       5658 CCGGCGCGGGTCCGTGCGCTCTTCGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 5717

Qy       3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
Db       5718 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCTCGTGTCGCCGATGCGTTCCGGAGCATG 5777

Qy       3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
Db       5778 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 5837

Qy       3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGACCTGCTCGACAGG 3900
Db       5838 CGTATTCCGACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGATCTGCTCGACAGG 5897

Qy       3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
Db       5898 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 5957

Qy       3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
Db       5958 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 6017

Qy       4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
Db       6018 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 6077

Qy       4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
Db       6078 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGACCCAAAAC 6137

Qy       4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
Db       6138 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 6197

Qy       4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
Db       6198 GCAGGCGTGCAAAGCGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 6257

Qy       4261 CTATGA 4266
Db       6258 CTATGA 6263
```

```
RESULT 9
AR172664
LOCUS       AR172664              71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1..71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others
ORIGIN

  Query Match             98.2%;  Score 4187.6;  DB 6;  Length 71989;
  Best Local Similarity   98.9%;  Pred. No. 0;
  Matches 4217;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
Db       1998 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGGGC 2057

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
Db       2058 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 2117

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
Db       2118 CGCGACACCGTCGGGCAAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 2177

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
Db       2178 GACCTCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 2237

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
Db       2238 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 2297

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
Db       2298 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 2357

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
Db       2358 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 2417

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 480
Db       2418 GCGCTGCCGCGAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATG 2477

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
Db       2478 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGTCCTCGGGCTGCGAGGGCCGTGTGTCGCG 2537
```

```
Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       2538 GTGGATACGGCCTATTCGTCCTCGCTCGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 2597

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2598 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 2657

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
              ||||||||||||||||||||||| |||||||||| |||||||||||||||||| ||||||
Db       2658 CTCGTGTGGCTCTCGAAGACCCGCGCGCTGGCCACGGACGGTCGCTGCAAGGCGTTTTCG 2717

Qy        721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2718 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 2777

Qy        781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       2778 AGTGGAGCCCGCGCGGACGGCGACCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 2837

Qy        841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2838 CACGACGGAGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 2897

Qy        901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2898 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 2957

Qy        961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2958 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 3017

Qy       1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       3018 CTCGGGCGAGACGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 3077

Qy       1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3078 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 3137

Qy       1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3138 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 3197

Qy       1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3198 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 3257

Qy       1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3258 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 3317

Qy       1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db       3318 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGGCCGGCAGAGCTGCTGGTGCTGTCGGCA 3377

Qy       1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
              ||||||||| ||||| |||||||||| |||||||||||||||||||||||||||||||||
Db       3378 AGGACCGCGGCAGCCTTGGATGCACACGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 3437

Qy       1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3438 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 3497

Qy       1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       3498 CACCGGCTCGCGGTGGCGGCGACGTCGAGCGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 3557

Qy       1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
              ||||||||||||||| ||||||||| ||||||| ||||||||||||||||||||||||||
Db       3558 GCGCAGGGACAGACGCCGCCCGGTGTGGTGCGCGGTATCGCCGATTCCTCACGCGGCAAG 3617

Qy       1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGCCGTGGCTGTAC 1680
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3618 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGCCGTGGCTGTAT 3677

Qy       1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3678 GATGTATGGCCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 3737

Qy       1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3738 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 3797

Qy       1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
              ||||||||||||||||| |||||||||||||||||||||||||| |||||||||||||||
Db       3798 CTCGACCAGACAGCCTTTACCCAGCCGGCGCTGTTCACCTTCGAGTATGCGCTCGCCGCG 3857

Qy       1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
              ||||||||||||||||| |||||||||||||||||||| |||||||||||||||||||||
Db       3858 CTGTGGCGGTCGTGGGGCGTAGAGCCGGAGTTGGTCGCTGGCCATAGCATCGGTGAGCTG 3917

Qy       1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3918 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 3977

Qy       1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       3978 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGCGGCGCCG 4037

Qy       2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4038 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 4097

Qy       2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4098 AACGGTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 4157

Qy       2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4158 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 4217

Qy       2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4218 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 4277

Qy       2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGC 2340
              |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db       4278 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTGGCACAGACGAGGTGAGC 4337

Qy       2341 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4338 TCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAG 4397

Qy       2401 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4398 GCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTC 4457

Qy       2461 GGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 2520
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       4458 GGCCTGGTGCCTGCCTGCCTGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCT 4517

Qy       2521 GGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 2580
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       4518 GGGCGTGACGAGCCAGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGC 4577

Qy       2581 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4578 CTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTAC 4637

Qy       2641 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4638 CCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGAC 4697

Qy       2701 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGC 2760
```

```
Db      4698 CGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGAAGGGGGGCGCGGTGCGCGGCGGC 4757

Qy      2761 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 2820
Db      4758 GACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCGGACGCCGGGAGAAGGTC 4817

Qy      2821 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTTGATCAC 2880
Db      4818 GAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGCCAGGCGTGCTCGATCGC 4877

Qy      2881 CTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC 2940
Db      4878 CTGGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTTGGCGAGGTCGAGATCGCCGTC 4937

Qy      2941 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 3000
Db      4938 GACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGGGCATGGTGCCCGACGAC 4997

Qy      3001 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 3060
Db      4998 CTGCCGGGAAAGCCCAACCCTCCGCTGCTGCTCGGAGGCGAGTGCGCCGGGCGCATCGTC 5057

Qy      3061 GCCGTGGGCGAGGGCGTGAACGGCCTCGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 3120
Db      5058 GCCGTGGGCGAGGGCGTGAACGGCCTTGTGGTGGGCCAACCGGTCATCGCCCTTTCGGCG 5117

Qy      3121 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 3180
Db      5118 GGAGCGTTTGCTACCCACGTCACCACGTCGGCTGCGCTGGTGCTGCCTCGGCCTCAGGCG 5177

Qy      3181 CTCTCGGCGATCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 3240
Db      5178 CTCTCGGCGACCGAGGCGGCCGCCATGCCCGTCGCGTACCTGACGGCATGGTACGCGCTC 5237

Qy      3241 GACAGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCATGCGGCGACCGGCGGG 3300
Db      5238 GACGGAATAGCCCGCCTTCAGCCGGGGGAGCGGGTGCTGATCCACGCGGCGACCGGCGGG 5297

Qy      3301 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 3360
Db      5298 GTCGGTCTCGCCGCGGTGCAGTGGGCGCAGCACGTGGGAGCCGAGGTCCATGCGACGGCC 5357

Qy      3361 GGCACGCCCGAGAAACGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 3420
Db      5358 GGCACGCCCGAGAAGCGCGCCTACCTGGAGTCGCTGGGCGTGCGGTATGTGAGCGATTCC 5417

Qy      3421 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 3480
Db      5418 CGCTCGGACCGGTTCGTCGCCGACGTGCGCGCGTGGACGGGCGGCGAGGGAGTAGACGTC 5477

Qy      3481 GTGCTCAACTCGCTCTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 3540
Db      5478 GTGCTCAACTCGCTTTCGGGCGAGCTGATCGACAAGAGTTTCAATCTCCTGCGATCGCAC 5537

Qy      3541 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 3600
Db      5538 GGCCGGTTTGTGGAGCTCGGCAAGCGCGACTGTTACGCGGATAACCAGCTCGGGCTGCGG 5597

Qy      3601 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 3660
Db      5598 CCGTTCCTGCGCAATCTCTCCTTCTCGCTGGTGGATCTCCGGGGGATGATGCTCGAGCGG 5657

Qy      3661 CCGGCGCGGGTCCGTGCGCTCTTGGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 3720
Db      5658 CCGGCGCGGGTCCGTGCGCTCTTCGAGGAGCTCCTCGGCCTGATCGCGGCAGGCGTGTTC 5717

Qy      3721 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCCCGTGTCGCCGATGCGTTCCGGAGCATG 3780
Db      5718 ACCCCTCCCCCCATCGCGACGCTCCCGATCGCTCGTGTCGCCGATGCGTTCCGGAGCATG 5777

Qy      3781 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 3840
Db      5778 GCGCAGGCGCAGCATCTTGGGAAGCTCGTACTCACGCTGGGTGACCCGGAGGTCCAGATC 5837

Qy      3841 CGTATTCCAACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGACCTGCTCGACAGG 3900
Db      5838 CGTATTCCGACCCACGCAGGCGCCGGCCCGTCCACCGGGGATCGGGATCTGCTCGACAGG 5897

Qy      3901 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 3960
Db      5898 CTCGCGTCAGCTGCGCCGGCCGCGCGCGCGGCGGCGCTGGAGGCGTTCCTCCGTACGCAG 5957

Qy      3961 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 4020
Db      5958 GTCTCGCAGGTGCTGCGCACGCCCGAAATCAAGGTCGGCGCGGAGGCGCTGTTCACCCGC 6017

Qy      4021 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 4080
Db      6018 CTCGGCATGGACTCGCTCATGGCCGTGGAGCTGCGCAATCGTATCGAGGCGAGCCTCAAG 6077

Qy      4081 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGGCCCAAAAC 4140
Db      6078 CTGAAGCTGTCGACGACGTTCCTGTCCACGTCCCCCAATATCGCCTTGTTGACCCAAAAC 6137

Qy      4141 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 4200
Db      6138 CTGTTGGATGCTCTCGCCACAGCTCTCTCCTTGGAGCGGGTGGCGGCGGAGAACCTACGG 6197

Qy      4201 GCAGGCGTGCAAAACGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 4260
Db      6198 GCAGGCGTGCAAAGCGACTTCGTCTCATCGGGCGCAGATCAAGACTGGGAAATCATTGCC 6257

Qy      4261 CTATGA 4266
Db      6258 CTATGA 6263
```

```
RESULT 10
AX024277
LOCUS       AX024277                16124 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024277 AX024384
VERSION     AX024277.1  GI:10184551
KEYWORDS    .
SOURCE      Polyangium cellulosum
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 82 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
COMMENT     On Oct 15, 2002 this sequence version replaced gi:10184588.
FEATURES             Location/Qualifiers
     source          1..16124
                     /organism="Polyangium cellulosum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match            30.2%;  Score 1287.4;  DB 6;  Length 16124;
  Best Local Similarity  72.5%;  Pred. No. 3.8e-131;
  Matches 1666;  Conservative    0;  Mismatches  631;  Indels    0;  Gaps    0;

Qy       444 CGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAAT 503
Db       134 CGAGGAGCAGGACGCATACGACATCACCGGCAATACGCTCAGCGTCGCCGACGGACGGTT 193

Qy       504 CTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTC 563
Db       194 GTCTTATACGCTAGGGCTGCAGGGACCCTGCCTGACCGTCGACACGGTCTGCTCGTCGTC 253

Qy       564 GCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCT 623
```

```
Db    254 GCTCGTGGCCATCCACCTTGCCTGCCGCAGCCTGCGCGCTCGCGAGAGCGATCTCGCGCT 313

Qy    624 GGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCG 683
Db    314 GGCGGGAGGCGTCAACATGCTCCTTTCGTCCAAGACGATGATAATGCTGGGGCGCATCCA 373

Qy    684 GGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACG 743
Db    374 GGCGCTGTCGCCCGATGGCCACTGCCGGACATTCGACGCCTCGGCCAACGGGTTCGTCCG 433

Qy    744 AGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGA 803
Db    434 TGGGGAGGGCTGCGGTATGGTCGTGCTCAAACGGCTCTCCGACGCCCAGCGACACGGCGA 493

Qy    804 TCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCT 863
Db    494 TCGGATCTGGGCTCTGATCCGGGGTTCGGCCATGAATCAGGATGGCCGGTCGACAGGGTT 553

Qy    864 GACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGG 923
Db    554 GATGGCACCCAATGTGCTCGCTCAGGAGGCGCTCTTGCGCGAGGCGCTGCAGAGCGCTCG 613

Qy    924 CTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGA 983
Db    614 CGTCGACGCCGGGGCCATCGGTTATGTCGAGACCCACGGAACGGGGACCTCGCTCGGCGA 673

Qy    984 CCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCC 1043
Db    674 CCCGATCGAGGTCGAGGCGCTGCGTGCCGTGTTGGGGCCGGCGCGGGCCGATGGGAGCCG 733

Qy   1044 GCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCAC 1103
Db    734 CTGCGTGCTGGGCGCAGTGAAGACAAACCTCGGCCACCTGGAGGGCGCTGCAGGCGTGGC 793

Qy   1104 TGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGC 1163
Db    794 GGGTTTGATCAAGGCGGCGCTGGCTCTGCACCACGAACTGATCCCGCGAAACCTCCATTT 853

Qy   1164 GCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCG 1223
Db    854 CCACACGCTCAATCCGCGGATCCGGATCGAGGGGACCGCGCTCGCGCTGGCGACGGAGCC 913

Qy   1224 GACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGAG 1283
Db    914 GGTGCCGTGGCCGCGGGCGGGCCGACCGCGCTTCGCGGGGGTGAGCGCGTTCGGCCTCAG 973

Qy   1284 CGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCGGC 1343
Db    974 CGGCACCAACGTCCATGTCGTGCTGGAGGAGGCGCCGGCCACGGTGCTCGCACCGGCGAC 1033

Qy   1344 GCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGGATGC 1403
Db   1034 GCCGGGGCGCTCAGCGGAGCTTTTGGTGCTGTCGGCGAAGAGCGCCGCCGCGCTGGACGC 1093

Qy   1404 ACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCGATGT 1463
Db   1094 ACAGGCGGCGCGGCTCTCAGCGCACATCGCCGCGTACCCGGAGCAGGGTCTCGGAGACGT 1153

Qy   1464 GGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGGCGAC 1523
Db   1154 CGCGTTCAGCCTGGTATCGACGCGTAGCCCGATGGAGCACCGGCTCGCGGTGGCGGCGAC 1213

Qy   1524 GTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGCCCGG 1583
Db   1214 CTCGCGCGAGGCGCTGCGAAGCGCGCTGGAGGTTGCGGCGCAGGGGCAGACCCCGGCAGG 1273

Qy   1584 TGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAGCTCGCCTTTCTCTTCACCGGACA 1643
Db   1274 CGCGGCGCGCGGCAGGGCCGCTTCCTCGCCCGGCAAGCTCGCCTTCCTGTTCGCCGGGCA 1333

Qy   1644 GGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGTTCCGCGA 1703
Db   1334 GGGCGCGCAGGTGCCGGGCATGGGCCGTGGGTTGTGGGAGGCGTGGCCGGCGTTCCGCGA 1393

Qy   1704 GGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCCGCGAGGT 1763
Db   1394 GACCTTCGACCGGTGCGTCACGCTCTTCGACCGGGAGCTCCATCAGCCGCTCTGCGAGGT 1453

Qy   1764 GATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCTTCACCCA 1823
Db   1454 GATGTGGGCCGAGCCGGGCAGCAGCAGGTCGTCGTTGCTGGACCAGACGGCGTTCACCCA 1513

Qy   1824 GCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGA 1883
Db   1514 GCCGGCGCTCTTTGCGCTGGAGTACGCGCTGGCCGCGCTCTTCCGGTCGTGGGGCGTGGA 1573

Qy   1884 GCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGCGT 1943
Db   1574 GCCGGAGCTCGTCGCTGGCCATAGCCTCGGCGAGCTGGTGGCCGCCTGCGTGGCGGGTGT 1633

Qy   1944 GTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGCAGGCGCT 2003
Db   1634 GTTCTCCCTCGAGGACGCCGTGCGCTTGGTGGTCGCGCGCGGCCGGTTGATGCAGGCGCT 1693

Qy   2004 GCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTGCTGCGGT 2063
Db   1694 GCCGGCCGGCGGCGCGATGGTATCGATCGCCGCGCCGGAGGCCGACGTGGCTGCCGCGGT 1753

Qy   2064 GGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCCGGACCAGGTGGTCAT 2123
Db   1754 GGCGCCGCACGCAGCGTTGGTGTCGATCGCGGCAGTCAATGGGCCGGAGCAGGTGGTGAT 1813

Qy   2124 CGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCGGGGCGCG 2183
Db   1814 CGCGGGCGCCGAGAAATTCGTGCAGCAGATCGCGGCGGCGTTCGCGGCGCGGGGGGCGCG 1873

Qy   2184 AACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGATGCTGGA 2243
Db   1874 AACCAAACCGCTGCATGTCTCGCACGCGTTCCACTCGCCGCTCATGGATCCGATGCTGGA 1933

Qy   2244 GGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCCTGGTCAG 2303
Db   1934 GGCGTTCCGGCGGGTGACTGAGTCGGTGACGTACCGGCGGCCTTCGATCGCGCTGGTGAG 1993

Qy   2304 CAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATTGGGTGCGCCA 2363
Db   1994 CAACCTGAGCGGGAAGCCCTGCACCGATGAGGTGAGCGCGCCGGGTTACTGGGTGCGTCA 2053

Qy   2364 CGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGG 2423
Db   2054 CGCGCGAGAGGCGGTGCGCTTCGCGGACGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGG 2113

Qy   2424 CACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCTGGTGCCTGCCTGCATGCC 2483
Db   2114 CCTCTTCGTCGAGGTGGGGCCGAAGCCGACGCTGCTCGGCCTTGTGCCGGCCTGCCTGCC 2173

Qy   2484 GGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGT 2543
Db   2174 GGATGCCAGGCCGGTGCTGCTCCCAGCGTCGCGCGCCGGGCGTGACGAGGCTGCGAGCGC 2233

Qy   2544 GCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTT 2603
Db   2234 GCTAGAGGCGCTGGGTGGGTTCTGGGTCGTCGGTGGATCGGTCACCTGGTCGGGTGTCTT 2293

Qy   2604 CCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTG 2663
Db   2294 CCCTTCGGGCGGACGGCGGGTACCGCTGCCAACCTATCCCTGGCAGCGCGAGCGTTACTG 2353

Qy   2664 GATCGACACGAAAGCCGACGACGCGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCA 2723
Db   2354 GATCGAAGCGCCGGTCGATCGTGAGGCGGACGGCACCGGCCGTGCTCGGGCGGGGGGCCA 2413

Qy   2724 CGACGAGGTCGAGGAGG 2740
Db   2414 CCCCCTTCTGGGTGAAG 2430
```

```
RESULT 11
AX024276
LOCUS       AX024276               10910 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024276 AX024383
VERSION     AX024276.1  GI:10184550
KEYWORDS    .
SOURCE      Polyangium cellulosum
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 81 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
COMMENT     On Oct 15, 2002 this sequence version replaced gi:10184587.
FEATURES             Location/Qualifiers
     source          1..10910
                     /organism="Polyangium cellulosum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match             21.8%;  Score 928.8;  DB 6;  Length 10910;
  Best Local Similarity   60.5%;  Pred. No. 3.3e-92;
  Matches 1606;  Conservative    0;  Mismatches 1027;  Indels   21;  Gaps    4;

Qy         25 GCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGTGATC 84
Db       2244 GCGGCGGATGACGACATCGCCATCGTCGGTGCCGCCTGCCGGTTCCCAGGTGGGGATGAG 2303

Qy         85 GATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGTCCCC 144
Db       2304 GGCCTGGAGACATACTGGCGGCATCTGGCCGAGGGCATGGTGGTCAGCACCGAGGTGCCA 2363

Qy        145 GCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAAGACG 204
Db       2364 GCCGACCGGTGGCGCGCGGCGGACTGGTACGACCCCGATCCGGAGGTTCCGGGCCGGACC 2423

Qy        205 CCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTTCGGC 264
Db       2424 TATGTGGCCAAGGGTGCCTTCCTCCGCGATGTGCGCAGCTTGGATGCGGCGTTCTTCGCC 2483

Qy        265 ATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGTGTGC 324
Db       2484 ATTTCCCCTCGTGAGGCGATGAGCCTGGACCCGCAACAGCGGCTGTTGCTGGAGGTGAGC 2543

Qy        325 TGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAACGGGA 384
Db       2544 TGGGAGGCGATCGAGCGCGCTGGCCAGGACCCGATGGCGCTGCGCGAGAGCGCCACGGGC 2603

Qy        385 GTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGCGTCC 444
Db       2604 GTGTTCGTGGGCATGATCGGGAGCG---AGCACGCCGAGCGGGTGCAGGGCCTCGACGAC 2660

Qy        445 GCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCGAATC 504
Db       2661 GACGCGGCGTTGCTGTACGGCACCACCGGCAACCTGCTCAGCGTCGCCGCTGGACGGCTG 2720

Qy        505 TCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTCCTCG 564
Db       2721 TCGTTCTTCCTGGGTCTGCACGGCCCGACGATGACGGTGGACACCGCCTGCTCGTCGTCG 2780

Qy        565 CTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGCCCTG 624
Db       2781 CTGGTGGCGTTGCACCTCGCCTGCCAGAGCCTGCGATTGGGCGAGTGCGACCAGGCCCTG 2840

Qy        625 GCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGACCCGG 684
Db       2841 GCCGGCGGGTCCAGCGTGCTTTTGTCGCCGCGGTCATTCGTCGCGGCGTCGCGCATGCGT 2900

Qy        685 GCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGGACGA 744
Db       2901 TTGCTTTCGCCAGATGGGCGGTGCAAGACGTTCTCGGCCGCTGCAGACGGCTTTGCGCGG 2960

Qy        745 GGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGGCGAT 804
Db       2961 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGCTCCGTGACGCGCAGCGCGACCGCGAC 3020

Qy        805 CGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGGTCTG 864
Db       3021 CCCATCCTGGCGGTGGTCAGGAGCACGGCGATCAACCACGATGGCCCGAGCAGCGGGCTC 3080

Qy        865 ACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGCGGACGCAGGC 924
Db       3081 ACGGTGCCCAGCGGTCCTGCCCAGCAGGCGTTGCTACGCCAGGCGCTGGCGCAAGCGGGC 3140

Qy        925 TGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGGTGAC 984
Db       3141 GTGGCGCCGGCCGAGGTCGATTTCGTGGAGTGCCACGGGACGGGGACAGCGCTGGGTGAC 3200

Qy        985 CCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCACGCCG 1044
Db       3201 CCGATCGAGGTGCAGGCGCTGGGCGCGGTGTACGGGCGGGGCCGCCCCGCGGAGCGGCCG 3260

Qy       1045 CTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCACT 1104
Db       3261 CTCTGGCTGGGCGCTGTCAAGGCCAACCTCGGCCACCTGGAGGCCGCGGCGGGCTTGGCC 3320

Qy       1105 GGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCACGCG 1164
Db       3321 GGCGTGCTCAAGGTGCTCTTGGCGCTGGAGCACGAGCAGATTCCGGCTCAACCGGAGCTC 3380

Qy       1165 CAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGCCCGG 1224
Db       3381 GACGAGCTCAACCCGCACATCCCGTGGGCAGAGCTGCCAGTGGCCGTTGTCCGCAGGGCG 3440

Qy       1225 ACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCATGAGC 1284
Db       3441 GTCCCCTGGCCGCGCGGCGCGCGCCCGCGTCGTGCAGGCGTGAGCGCTTTCGGCCTGAGC 3500

Qy       1285 GGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCCGGCG 1344
Db       3501 GGGACCAACGCGCATGTGGTGTTGGAGGAGGCGCCGGCGGTGGAGCCTGTGGCCGCGGCC 3560

Qy       1345 CCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGGATGCA 1404
Db       3561 CCCGAGCGCGCAGCGGAGCTGTTCGTCCTGTCGGCGAAGAGCGCGGCGGCGCTGGATGCG 3620

Qy       1405 CAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCGATGTG 1464
Db       3621 CAGGCAGCCCGGCTGCGGGACCACCTGGAGAAGCATGTCGAGCTTGGCCTCGGCGATGTG 3680

Qy       1465 GCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGGCGACG 1524
Db       3681 GCGTTCAGCCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTGGCGGTGGCCGCGAGC 3740

Qy       1525 TCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGCCCGGT 1584
Db       3741 TCGCGCGAGGCGCTGCGAGGGGCGCTTTCGGCCGCAGCGCAGGGGCACACGCCGCCGGGA 3800

Qy       1585 GCGGTGCGCAGTATCGCCGATTCCTCACGCG---GCAAGCTCGCCTTTCTCTTCACCGGA 1641
Db       3801 GCCGTGCGTGGGCGGCCTCGGGCGGCAGCGCGCCGAAGGTGGTCTTCGTGTTTCCCGGC 3860

Qy       1642 CAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGTTCCGC 1701
Db       3861 CAGGGCTCGCAGTGGGTGGGCATGGGCCGAAAGCTCATGGCCGAAGAGCCGGTCTTCCGG 3920

Qy       1702 GAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCCGCGAG 1761
Db       3921 GCGGCGCTGGAGGGTTGCGACCGGGCCATCGAGGCGGAAGCGGGCTGGTCGCTGCTCGGG 3980

Qy       1762 GTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCTTCACC 1821
```

```
Db      3981 GAGCTCTCCGCCGA---------CGAGGCCGCCTCGCAGCTCGGGCGCATCGACGTGGTT 4031

Qy      1822 CAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTA 1881

Db      4032 CAGCCGGTGCTGTTCGCCATGGAAGTAGCGCTTTCTGCGCTGTGGCGGTCGTGGGGAGTG 4091

Qy      1882 GAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGGCGGGC 1941

Db      4092 GAGCCGGAAGCGGTGGTGGGCCACAGCATGGGCGAGGTTGCGGCGGCGCACGTGGCCGGC 4151

Qy      1942 GTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGCAGGCG 2001

Db      4152 GCGCTGTCGCTCGAGGACGCGGTGGCGATCATCTGCCGGCGCAGCCGGCTGCTGCGGCGG 4211

Qy      2002 CTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTGCTGCG 2061

Db      4212 ATCAGCGGTCAGGGGGAGATGGCGCTGGTCGAGCTGTCGCTGGAGGAGGCCGAGGCGGCG 4271

Qy      2062 GTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCCGGACCAGGTGGTC 2121

Db      4272 CTGCGTGGCCATGAGGGTCGGCTGAGCGTGGCGGTGAGCAACAGCCCGCGCTCGACCGTG 4331

Qy      2122 ATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCGGGGCG 2181

Db      4332 CTCGCCGGCGAGCCGGCGGCGCTCTCGGAGGTGCTGGCGGCGCTGACGGCCAAGGGGGTG 4391

Qy      2182 CGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGATGCTG 2241

Db      4392 TTCTGGCGGCAGGTGAAGGTGGACGTCGCCAGCCATAGCCCGCAGGTCGACCCGCTGCGC 4451

Qy      2242 GAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCCTGGTC 2301

Db      4452 GAAGAGCTGATCGCGGCGCTGGGAGCGATCCGGCCGCGAGCGGCTGCGGTGCCGATGCGC 4511

Qy      2302 AGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATTGGGTGCGC 2361

Db      4512 TCGACGGTGACGGGCGGGGTGATCGCGGGTCCGGAGCTCGGTGCGAGCTACTGGGCGGAC 4571

Qy      2362 CACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGGCCGGTGCG 2421

Db      4572 AACCTTCGGCAGCCGGTGCGCTTCGCTGCGGCGGCGCAAGCGCTGCTGGAGGGTGGCCCC 4631

Qy      2422 GGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCTGGTGCCTGCCTGCATG 2481

Db      4632 GCGCTGTTCATCGAGATGAGCCCGCACCCGATCCTGGTGCCGCCCCTGGACGAGATCCAG 4691

Qy      2482 CCGGACGCC------CGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGACGAGCCG 2535

Db      4692 ACGGCGGCCGAGCAAGGGGGCGCTGCGGTGGGCTCGCTGCGGCGAGGGCAGGACGAGCGC 4751

Qy      2536 GCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCCTGGGCC 2595

Db      4752 GCGACGCTGCTGGAGGCGCTGGGGACGCTGTGGGCGTCCGGCTATCCGGTGAGCTGGGCT 4811

Qy      2596 GGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGCAGCGCGAG 2655

Db      4812 CGGCTGTTCCCCGCGGGCGGCAGGCGGGTTCCGCTGCCGACCTATCCCTGGCAGCACGAG 4871

Qy      2656 CGCTACTGGATCGA 2669

Db      4872 CGGTACTGGATCGA 4885
```

```
RESULT 12
POL505006
LOCUS       POL505006             48000 bp    DNA     linear   BCT 06-JAN-2003
DEFINITION  Sorangium cellulosum spiG gene (partial), spiH gene, spiI gene,
            spiJ gene, spiK gene, spiL gene and spiZ gene.
ACCESSION   AJ505006
VERSION     AJ505006.1  GI:27528333
KEYWORDS    cytochrome-P450-monooxygenase; OH-methyltransferase; polyketide
            synthase; regulatory protein; spiG gene; spiH gene; spiI gene; spiJ
            gene; spiK gene; spiL gene; spiZ gene.
SOURCE      Polyangium cellulosum
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1
  AUTHORS   Knauber,J., Bloecker,H., Mueller,R., Nordsiek,G. and Beyer,S.
  TITLE     The spirangiene synthase from Sorangium cellulosum So ce90 - Module
            duplications as strategy behind the evolution of polyketide
            synthases
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 48000)
  AUTHORS   Beyer,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2002) Beyer S., Nbi-Mx, German Research Centre
            for Biotechnology, Mascheroder Weg 1, D-38124 Braunschweig, GERMANY
FEATURES             Location/Qualifiers
     source          1..48000
                     /organism="Polyangium cellulosum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:56"
     gene            1..7558
                     /gene="spiG"
     CDS             <1..7558
                     /gene="spiG"
                     /function="biosynthesis of spirangiene"
                     /codon_start=2
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="CAD43448.1"
                     /db_xref="GI:27528334"
                     /translation="ILGSDARLASALEAAGAHVTRHRDPGAWIDAIRGGGAPEVVAVL
                     ACTTRPDAPSGALAATHGAAADALAALQAWLAEERLASCALVVLTERAVATRPDEDVL
                     DLVHAPVWGLVRAAQVENPDRALILVDVDADDASLPALPALVAAAVARAADASEPQLA
                     LREGRRLVPRLGRLGGDDALVPPSAAAYRLHVAAKGTLESLALVPHEEATRPLAEGQI
                     RVAVRAAGLNFRDALDALGMYPGDPGPLGAEGAGVVLEVGPGVTELAPGDRVMGLLSA
                     AFGPIALADHRMVSRIPAGLSFVEAASIPVVFLTAYYGLVDLGRLRPGERVLVHAAAG
                     GVGIAAVQLARHLGAEVFATASPPKWATLRALGLDDDHIASSRSGDFEAHFRRTSGGR
                     GVDVVLDSLARELVDASLRLVALAPDGRFVEMGKSDVRDPLAVAASHPSVAYRAFDLA
                     EAGPERIQAMLAELVALFERGVLRVPPITPCDLRRAPQAFRALAQARLVGKVVLTIPR
                     ALDPRGSVLVTGGTGTLGALVARHLVQRHGVRSLVLSSRRGPAAPGADDLRRELEAAG
                     ASVAIVACDAGDRLALERLLAGIPADRPLTAVVHAAGTLDDGVLTSLTPERLRSVMSA
                     KVDAAAHLDELTRPLDLSAFVLFASLSGVLGSPGQANYAAANTFLDALAHQRRSRGLP
                     AVSLDWGYWAERTGLTAHLGDTDLQRMARGGLRPLTTDDGLALFDAALAHVDAALVAA
                     PFDPVALRAHAVAPILRGLVRAGAARRKLANNAASAAPLAQRLAGLAPKERERAMIDV
                     VRPEIGVVLGLPPSRHLDGDRPLKELGLDSLMAVELRNRLAAIVGTRLPATLLFDYPT
                     PTALARFLATALVGRGDAAAAPSAARAASAGASDDAIAIVAMACRFPGGVLTPEDLFR
                     LLREGGDAISAFPENRGWDLDALSDGAARQGGFLHDADLFDAEFFGISPRETLALDPQ
                     QRLLLEMTWEALERAGIIPASLHGSKSGVFVGVFGSDYVARLLHGPSGAQAMESLRGY
                     IGTGSASSVASGRIAYALGLEGPAISVDTACSSSLVALHLACHAIRRGECSLALAGGV
                     TVMATPGIFGEFSAQHAGAPDGRCKSFSADANGAGWSEGAGIVLLERLSDARKNGHPV
                     LAIVRGSAVNQDGKSQGLTAPNGPAQERVIRDALASAGLAPADVDAVEAHGTGTTLGD
                     PIEARALFATYGAAHAKERPLWLGSLKSNLGHTQAAAGIGGVIKMVLALQSGVLPRTL
                     HAASPSQHIDWSSGAIRLLTEEVPWKPGGSPRRAAVSSFGISGTNAHVVLEEAPRAAA
                     GDDQAAEPARAPLAPPVPVLLSAKSDAALRAQAGKLHAHLAEHEDVALVDVAASLATT
                     RSHFERRAAVVADDRAALLSALAALAEGRAGAGTVAGEALPPGKVAFVFPGQGSQWPS
                     MARALLASSPAFRAEIEACERALAPHVDWSLLAVLGGDEAHAAPMLERVDVVQPVLFA
                     VMIALAATWRAAGVTPDAVVGHSQGEIAAAYVAGALSLEDAARVVALRSRAITKLAGR
                     GAMAAVELTTADLEARLAPFGGRLAIAAINSPHAALVSGDPDAIDELVAELSGAQLFA
                     RKVRVDYASHSAHVEAIERTLLEALDGIAPRPATVPLYSAVTGERLDGEALGAAHWYR
                     NLRHTVRFEHATRALLDDGHRFFVEVSPHPVLTVALRDTIDASGHDAVACASLRRDEG
                     DLARFNLSLAELHTRGLAVDWRAFFAPFAPRTVALPTYAFQRERFWLDAPKASGVIPA
                     AASPAEARFWSAVEDGDVATVTDALDVGDDERRAALATLLPTLATWRRARRAQSTIDA
                     WRYRVAWRALTAPPKGDIAGTWLVVAPEVAGALADGVARALEARGARGVLVPVDAEDA
                     ARTPLAGRLRAVAGGQAIRGVISFAALDERPAPLEPSVPLGLAVSLGLVQALGDAAID
                     APLWLVTRGAVAVGASERVTRPLQAMTWGLGRVVALEHPERWGGLVDLGEESAPDANG
                     ARDANVHDRLANLVARRDDEDQLALRAAGTYVRRLVRAPHDGARPPAFAPRGAALVTG
                     GTGALGAHVARWLARIGAEHIVLASRRGAAAPGAAALAEELSALGARVTLAACDVADR
                     EAVAALVRDVEAGGATVRAVFHAGGAMHEAPVAAMRVEELADAIAAKARGAQHLHDVF
                     AQRPLDAFVLFSSGAGVWGGGRQGAYAAANAFLDALAEARRADGLAATSIAWGAWAGG
                     GMLATDAERRLKHRGVAPMDPELAVAALAHALDHAETCLAVADVDWARFAPSFASARP
                     RPLLDELAEARSALDALREPPDDARTAAGPEPASTLRTTLAALPEGERHRHLLALVRT
                     ETAAVLGHADASRVEPNRGFFDLGLDSLMSVELRRRVQRATGIKLPATLAFDHPTPSA
```

```
                  LASKVLAAIVLHDATPRASPAAELERLEGMLSAIYADEALRDDLTARLRAFLDKRAVR
                  TERPDDAAFAEKLGSASADELIRLIDQKLGDRIDVDRY"
gene              7542. .22595
                  /gene="spiH"
CDS               7542. .22595
                  /gene="spiH"
                  /function="biosynthesis of spirangiene"
                  /codon_start=1
                  /evidence=experimental
                  /transl_table=11
                  /product="polyketide synthase"
                  /protein_id="CAD43449.1"
                  /db_xref="GI:27528335"
                  /translation="MSTVTNDTLTEYLRRLTQELHRSETRLRATEERRHEPIAIVGLG
                  LRLPGGIHDRDTLWTFLEEGRDAIAPIPASRWNADATYDPDPDAVGKSYVRDAAMLDR
                  VDLFDADFFGISPREAKYVDPQHRLLLETSWQALEDAGIVPASLRDSKTGVFVGTGAS
                  DYAFLQSDRDASEAYAFMGMISSFAAGRLAFTLGLQGPALSIDTACSSSLVALHLACQ
                  SLRQGECDLALVAGVQVMSSPEVFVLLSRTRALASDGRSKTFSANADGYGRGEGVVVL
                  AVERLRDARAKGRPILAVIRGSAVNHDGTSSGITVPNGPAQQKVLRAALDDARLVPAD
                  VDVVECHGTGTSIGDPIEVNALAAVYGEGRPKDRPLFLGALKTNIGHLEFASGLAGVA
                  KMVASMRHATLPATLHTSPLNPLVDWDALPVRVVDAARPWTRRDDGAPRRAGVTAIGL
                  SGTNAHVIVEEAPAEPEPTTPDAAPALPAVPVLLSGKTDEALRAQAARLHAHLAGRPD
                  ARLVDIAASLATTRTHFDRRAAVVAADRDELLGALDALARGEAGPGSVVASAIPAGRV
                  VFVFPGQGSQWVGMARALLASSVVFRDEIAACERALAPHVAWSLGAVLRGDGDEATLL
                  GRVDVVQPVLFAVMVALAALWRSIGVTPDAVVGHSQGEIAAAYVAGALSLEDAAKVVA
                  LRARALTKIAGRGAMAAVELGARDTEARLAPFGDAIAIAAINSPRATLVAGDTDAIDA
                  LVRDLEAAQIFARKVRVDYASHSAHVEAIERELLADLAGIEPRAGAVPLYSAVTGAKL
                  DGSRLDAAHWFRNLRSTVRFEDATRALHDDGHRFFVEVSPHPVLALALDETLSSIGGD
                  AAVVSSLRRDEGDLERFLLSAGELHARGHAIDWGAFFAPLGARRVTLPTYAFQRERFW
                  LDAGDASDDEDAAPGASAEETAFWRAIERGDVAALSDALHVADSGRRSALESLMPALS
                  AWRRSRREQSTVGAWRYRVEWRPVSAASRGDVAGTWLVVAPAGVASDLASALAAALTE
                  RGADVVPLALEASDITAHDAASRLRDAIAGRPAIRGVLTLTPLDTAPLDGRPAVPAGL
                  ARSLALVQALAELAIEAPIWIVTRGAVAIGRSDRIESPVQAMAWGLGRVVSLEQPERW
                  GGLVDLPRAHADGALDATAIERLVAALGSREGEDELALRPTGLFARRLARAPLGDAPA
                  QRFEPRGTVLVTGGTGALGAHVARWLARHGAAHLVLASRRGAAAPGAEALRAELTELG
                  ARVTIAACDVADRAAIEELFAALPAEEPLTAIFHTAGVIDDGVLGSLTPERLETVLHG
                  KADAARTLHELTKTRDLAAFVLFSSAAGVFGNAGQASYAAANAFLDALAEERRALGLR
                  ATSVAWGAWAGAGMLTFDVAQGLRRGGFAAMEPEIAVAALASALDRDETHVTIASIDW
                  ARFAPSYASARQRPLVADLPDARRALDAQPSASTEHDETALVATLRALSGEARLRHLV
                  SLVLAQTAAVLGHADRSRLDPQRGFFDLGLDSLMAVELRRRLQKATGVKLPATITFDR
                  PSPHRLAGFLLDALPLGGAAPSTDARPRAARADASDEPIAIVGVGLRLPGGVDDVDAL
                  FRLLEGERDAVAPIPASRWDVGAVFDPDPEAKGKTYVRHAAMLDRVDLFDAGFFGISP
                  REARHVDPQHRLLLETAWQALEAATIVPASLRDSTTGVFVGAGASDYAVLQSSAEDAE
                  AYAAMGTAASFAAGRLAFTLGLQGPALSIDTACSSSLVALHLACQSLRQGECDLALAA
                  GVQVMASPEVFVILSRTRALAPDGRSKTFSANADGFGRGEGVVVLALERLRDARANGH
                  PVLAVVRGTAVNHDGASSGITAPNGSSQQKVLRAALDDARLAPTDVDVVECHGTGTSL
                  GDPIEVQALAAVYGEGRAPERPLRLGAVKTNIGHLEFASGLAGVVKIVASLRHGTLPA
                  TIHTNPRNPHIDWDALPVRVVDAAEPWERSADGSPRRAGVSAFGLSGTNAHVIVEEAP
                  EPDAPKAARGALPLSGALPFVLSAKSDAALRAQAAALRDQLARTPDAALVDVAAALAT
                  TRSQFDHRAAIAASDHGALVAALEGLAAGVAAPGTFVAKGAADKLAFLFTGQGAQRAA
                  MGRGLYDAFPVFRDALDAVASHLDRELDRPLRDVLFAPKGSELASLLDRTEFTQPALF
                  ALEVALFRLVEAWGVTPDVLLGHSVGELAAAHVAGVLSLADACTLVAARARLMQALPA
                  RDGAMVTVHATEAEVLAALEGRDGRAEIAAINAPSSTVIAGDVDAVLRVAAHFEARGR
                  KATRLRVSHAFHSHHMDPMLDAFRRVAEGLTFHPPRIPIVSNVTGRLAADEIRSPDYW
                  VRHVRSAVRFADGINTLEADGVSSFFELGPHGVLSALGESALANAQREVAFVPALRDG
                  RADVDALTAALSSLHVRGHRVDWAAFFGPFDPQTVALPTYAFQRDRFWLDAPAGATRD
                  EAAPAAPEDAAFWRAVDAGDVGALGATLNASGEDHLSALATLLPALSAWRRARDERSL
                  VDALRYRVVWKPLTPTAARDVAGTWLLVTLAGAEDALARELERALTARGAEVITMPIA
                  PAEADRVRLAVRVQGALADAGELRGIVSFAALDETTLATHTALPAGLALTLRLVQALG
                  DMGIEAPLWLVTRGAVSTGRSDRLASAAQSMTWGLGRVVGLEHPERWGGLVDVADGDT
                  IDARALDRLVTLLASRDGEDQLALRPTGIFARRLVRAPLGDAPPARTYTARGTSLVTG
                  GTGALGAHVARFLAGRGAEHLVLVSRRGEDAPGAGALRAELEALGARVTLAACDVADR
                  ASVAALLARLDAQGDPVRAVVHAGGVVAQAALASTDLDGVQAVVAAKVRGALHLHELL
                  GDRELDAFVLFASGAGVWGSGQQGAYAAGNAFLDALAEVRRAAGLTATSIAWGAWAGG
                  GMLADHADADAEVELRKRGLVPMAPALAIAALAGALDHGETAVTVASVDWARFAPAFA
                  SARPRPLLLDLPEAQRAIEAPSAPAGADGASADRALVEALRPLSEADRARHLVELVVA
                  ETAAVLGHADASAIDPHKGFFDLGLDSLVAVELLRKLRARTGVELPATVTFDHPSPHR
                  LATYVREALPLDAPLAPADATTARSSSAHAAIDEPIAIVGLGLRLPGGVEDREGLWSF
                  LEQGRDGVGPIPKDRWDADAIYDPDPEAKGKSYVREAALLDRVDLFDAAFFGISPREA
                  KQVDPQHRLLLETAWQALEEATIVPASLKDTSTGVFVGIGLSDYALMQGSAEETDAYA
                  AMGTLSSFAAGRLAFTLGLQGPALSIDTACSSSLVALHLACQSLRKGECDLALAGGVQ
                  VMTLAEPFVLLSRTRALAADGRSKTFSAHADGYGRGEGVVVLALERLRDAEARGRRVL
                  AVVRGTAVNHDGASSGITAPNGTSQQKVLRAALGDAGLAPADVDVVECHGTGTALGDP
```

```
                  IEVQALAAVYGEGRAAERPLLLGAVKSNIGHLEAASGLAGVAKIVTALEHGALPATIH
                  SRPRNPHVDWDALHVRVVDAAELWARRDEETPRRAGVSAFGLSGTNAHAIIEEAPRAR
                  RGAAATDEPPPAVPVVLSAKSERALRAQAERLRAHVTGRPGDALVDVAASLATTRSHF
                  EHRAAIVARDLGALEDALAGLAEGRSPAGTVVSRTTGGKLALLFTGQGGQRAAMGRGL
                  YGAFSVFRDAFDAACSYLDRELEHPLRDVVFAPRGSELAALLDETLYTQTSLFALEVA
                  LFRLVEAWGVTPHVLLGHSIGELVAAHVAGVLSLEDACTLVAARARLMHALPRRDAAM
                  VAVHACESEVLDALDAMAGLDGRASIAAVNAPSSTVVAGDREAVLAVAARFEALGRKT
                  SRLRVSHAFHSHHLDPMLEAFRRVAEGVRFEPPRIPIVSNVTGARASNDDLCSADYWV
                  RHARGTVRFADGLEALRADGVSTFLELGPHAVLCALGHETLESEAVAFLPALRDGRDD
                  VDTLTAALGALHARGHHVDWDAFFAPFNAGRVALPTYAFQRERFWLDMPKAKSSRAAA
                  ADSAGRYPLAGTRLDLPDGSAVHTLDIGPGAQPYLADHSVYGRIVVPGAFHLAVLLAV
                  AESHWPDSPLELRDVEFVRALTFEEPSETVAVRVHLTPAGGDGKGFHATVATQTEGAW
                  TTHATATLGAAPPSALARHPRAAHEPLAGDEPSRAVARLMASFEAFNVEWGPKWRWLR
                  QTTHAGDRTGVGYLAAPDGVPTDDAPLPGGLVDNAFAVVGSAVTGLPAHDNRSTSNVP
                  PLPFSIERLVWYGRAETPCWADFVMRGEITPDADRYIGDITLWSREGASSEGASREGT
                  PIATIEGLAFRRAPADKFLPELPARNLYALTWPEQPAASRPPRGRWALLGEDSLGIAT
                  GIAADAYPDVDALRDALGRGAPCPDTVAVCLATAHGPSDTDETELVSSVHTAAARALA
                  LLNAWLSDERLASSTLVLLTQRAVATCAGEDVLDLASAALWGLVRSAQAENRDRAILL
                  VDLDGSDASRRALPYAVGGALDDAEPQLALRDGRRLAPRLARVAAPTAATNAATSTS
                  ATPRALDPEGTVLVTGGTGTLGALVARRLVERHGVKHLVLLSRRGPDAPGATDLAAEL
                  QARGASVVLAAADAAEPAALARVLLAIPHDRPLTAVVHAAGTLDDGVLSSMTPARLSA
                  VLRAKVDAAVNLDEQTRHSPLRAFVLFSSLSGVLGSPAQSNYAAANAFLDALAHRRAA
                  RGLPAVALDWGYWAQTSGLTKHLEQADLRRIARGGLRPLASDEALALFDLALAQDAPC
                  LVPAAFDADALRKSGDALSALLRGLVAGPRARAQRASGAGAASALKLRLASLAPLERH
                  RMLLDVVRADVATVLGVASPLDIEPGRPLQELGLDSLMAIELRNRLAAATGLRLSATL
                  LFDHPTPEALAKLISGKLLHDEAAPPSTSAAEIDRLEGALASAYANEAIRDELTVRMK
                  AILSKWMAAREPSPDDGLASRLDTANDDELFRMIDHVRAEAAT"
gene              22592. .33133
                  /gene="spiI"
CDS               22592. .33133
                  /gene="spiI"
                  /function="biosynthesis of spirangiene"
                  /codon_start=1
                  /transl_table=11
                  /product="polyketide synthase"
                  /protein_id="CAD43450.1"
                  /db_xref="GI:27528336"
                  /translation="MSAPPGTEAKLRDYLKQLTLDLQRSQTRLHAAESKAYEPIAIVA
                  MSCRFPGGARSPEDLWQLLEGGRDAVSGFPASRGWSVEPRVDPRADVTGKTPALEGGF
                  LEGAELFDAAFFGVSPREALEMDPQQRLLLETTWEAIERAEIDPASLQGSQTGVFVGV
                  MYNNYASHHLGGDGDASGYVSFASSASVASGRIAYTFGFHGPTVTVDTACSSSLVAVH
                  LACQALRSGECALALAGGVTVMATREAFDALSFAGSPDGRCKSFSADANGAGWSEGAG
                  MVLLERLSDAQRNGHPILAVIRGSAVNQDGKSQGLTAPNGPAQERVIRQALAAARLET
                  HDIDAVEAHGTGTPLGDPIEANALLATYGLGRPAERPLWLGSLKSNVGHTQAAAGVGG
                  IIKTVLALRNGLLPRTLHAHAPSPHVDWAAGGLRLLADAVPWTRNGRVRRAAVSSFGL
                  SGTNAHVIVEEAPRPDRDEASPAPSGRDQALPVLLSAKTEDALRAQAARLHAHLAGDP
                  GASLVDVAASLATTRAHFDHRAAFVARSRDALLAALDAFARGAIDAPASVASVARRAP

  Query Match             20.1%;  Score 859;  DB 1;  Length 48000;
  Best Local Similarity   57.6%;  Pred. No. 7.7e-85;
  Matches 1735;  Conservative    0;  Mismatches 1230;  Indels   48;  Gaps    9;

Qy         14 CCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCG 73
              || ||  |  |||   ||| || ||  || ||||||||||       || || | |
Db      12286 CCCGCGCGGACGCGAGCGACGAGCCCATCGCGATCGTCGGCGTCGGCCTGCGCCTCCCAG 12345

Qy         74 GTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCG 133
              | || ||   |||  |    || | |||       || |||||||| |||||| | ||||
Db      12346 GCGGGGTCGACGACGTCGACGCGCTCTTCCGCCTGCTCGAGGGCGAGCGCGACGCGGTCG 12405

Qy        134 GGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCC 193
                |   |||| ||    |||||||| |   |   ||     ||| ||||| |||||||| ||
Db      12406 CTCCGATCCCGGCGAGCCGCTGGGACGTCGGCGCCGTGTTCGATCCGGACCCCGAAGCGA 12465

Qy        194 CGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCT 253
                ||||||||  |||     | |||   |   | ||    |  |||  | |    ||||·||||
Db      12466 AGGGGAAGACGTACGTCCGCCACGCGGCGATGCTCGACCGCGTCGACCTCTTCGACGCGG 12525

Qy        254 CCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGC 313
              ||||||||||||      ||  |||||  |||   ||      |  |||||      || || ||||||
Db      12526 GCTTCTTCGGCATCAGCCCGCGCGAGGCGAGGCACGTCGACCCGCAGCACCGCCTCTTGC 12585

Qy        314 TGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTA 373
```

```
Db      12586 TCGAGACCGCGTGGCAAGCGCTCGAGGCGGCGACGATCGTCCCCGCGTCGCTCCGCGACT 12645

Qy        374 CGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAG 433
Db      12646 CGACCACCGGCGTCTTCGTCGGCGCGGGCGCGAGCGACTACGCCG---------TGCTCC 12696

Qy        434 CGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAG 493
Db      12697 AGAGCAGCGCCGAGGACGCCGAGGCCTACGCCGCGATGGGCACCGCCGCCTCGTTCGCGG 12756

Qy        494 CGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCT 553
Db      12757 CAGGTCGGCTGGCGTTCACGCTCGGCCTCCAAGGGCCCGCGCTCTCGATCGACACCGCGT 12816

Qy        554 ATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCT 613
Db      12817 GCTCGTCGTCGCTCGTCGCGCTCCACCTCGCGTGCCAGTCGCTGCGGCAAGGCGAGTGCG 12876

Qy        614 CCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCT 673
Db      12877 ACCTCGCGCTCGCCGCCGGCGTCCAGGTGATGGCCTCGCCCGAGGTCTTCGTGATCCTCT 12936

Qy        674 CGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATG 733
Db      12937 CGCGCACGCGCGCCCTCGCTCCGGACGGGCGATCGAAGACGTTCTCGGCCAACGCCGACG 12996

Qy        734 GGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCG 793
Db      12997 GGTTCGGCCGCGGCGAAGGCGTCGTCGTCCTCGCCCTCGAGCGCCTCCGGGACGCGCGCG 13056

Qy        794 CGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGA 853
Db      13057 CCAACGGCCACCCGGTCCTCGCCGTCGTCCGCGGCACCGCGGTCAACCATGACGGCGCGT 13116

Qy        854 GCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGG 913
Db      13117 CGAGCGGCATCACCGCGCCGAACGGCTCCTCGCAGCAGAAGGTCCTCCGCGCCGCCCTCG 13176

Qy        914 CGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGA 973
Db      13177 ACGACGCCCGCCTCGCTCCCACCGACGTCGACGTCGTCGAGTGCCACGGCACCGGCACCT 13236

Qy        974 CGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATG 1033
Db      13237 CCCTCGGCGATCCCATCGAGGTCCAGGCCCTCGCGGCCGTCTACGGCGAAGGGCGCGCCC 13296

Qy       1034 TCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGT 1093
Db      13297 CCGAGCGCCCGCTCCGCCTCGGCGCGGTGAAGACCAACATCGGCCACCTCGAGTTCGCGT 13356

Qy       1094 CGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGC 1153
Db      13357 CGGGCCTCGCCGGCGTCGTCAAGATCGTGGCGTCGCTCCGCCACGGGACGCTCCCCGCGA 13416

Qy       1154 ACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCA 1213
Db      13417 CGATCCACACCAACCCGCGCAACCCGCACATCGACTGGGACGCGCTCCCCGTCCGCGTCG 13476

Qy       1214 CGCGCGCCCGGACACCGTGGCCG------GACTGGAATACGCCGCGACGGGCGGGGGTGA 1267
Db      13477 TCGACGCGGCCGAGCCGTGGGAGCGGAGCGCCGACGGATCTCCGAGGCGCGCGGGGGTCT 13536

Qy       1268 GCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGG----CG 1323
Db      13537 CCGCGTTCGGGCTCTCCGGCACGAACGCGCACGTCATCGTCGAAGAGGCGCCCGAACCCG 13596

Qy       1324 GCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGC--------TGGTGCTGT 1375
Db      13597 ACGCCCCGAAGGCCGCGCGCGGCGCGCTGCCCCTCTCGGGCGCGCTCCCCTTCGTGCTCT 13656

Qy       1376 CGGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGA 1435
Db      13657 CGGCCAAGAGCGACGCGGCCCTCCGCGCGCAAGCCGCGGCGCTCCGCGATCAGCTCGCGA 13716

Qy       1436 CCTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGA 1495
Db      13717 GGACACCCGACGCGGCGCTCGTCGACGTCGCCGCGGCGCTCGCCACGACGCGCTCGCAAT 13776

Qy       1496 TGGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACG 1555
Db      13777 TTGACCACCGCGCCGCGATCGCCGCGAGCGATCACGGGGCGCTGGTCGCGGCGCTCGAGG 13836

Qy       1556 CTGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCG 1615
Db      13837 GCCTCGCCGCGGGCGTCGCGGCGCCAGGCACGTT------CGTCGCCAAGGGCGCCGCCG 13890

Qy       1616 GCAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGC 1675
Db      13891 ACAAGCTCGCCTTCCTGTTCACGGGCCAGGGCGCGCAGCGCGCTCGCGATGGGGCGCGGCC 13950

Qy       1676 TGTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACC 1735
Db      13951 TGTACGACGCGTTCCCCGTCTTCCGCGACGCGCTGGATGCCGTCGCGAGCCACCTCGATC 14010

Qy       1736 AGGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCG 1795
Db      14011 GCGAGCTCGATCGCCCGCTCCGCGACGTCCTCTTCGCGCCGAAGGGCTCGGAGCTCGCGA 14070

Qy       1796 CGCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCG 1855
Db      14071 GCCTCCTCGACCGGACCGAGTTCACGCAGCCGGCGCTCTTCGCCCTCGAAGTCGCGCTCT 14130

Qy       1856 CCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTG 1915
Db      14131 TCCGCCTCGTCGAGGCGTGGGGCGTCACGCCCGACGTGCTCCTGGGGCACTCCGTCGGCG 14190

Qy       1916 AGCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGG 1975
Db      14191 AGCTCGCCGCCGCGCACGTCGCGGGCGTCCTCTCGCTCGCCGACGCGTGCACGCTCGTCG 14250

Qy       1976 CTGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGC---CGGCGGGGCGATGGTGTCGATCG 2032
Db      14251 CCGCGCGCGCGCGCCTCATGCAGGCCCTGCCCGCGCGCGACGGCGCGATGGTCACGGTGC 14310

Qy       2033 AGGCGCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCG 2092
Db      14311 ACGCGACGGAGGCCGAGGTGCTCGCCGCGCTCGAGGGGCGGGACGGGCGCGCCGAGATCG 14370

Qy       2093 CCGCGGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGA 2152
Db      14371 CCGCGATCAACGCGCCGTCCTCGACGGTGATCGCCGGCGACGTCGACGCCGTGCTGCGCG 14430

Qy       2153 TCGCGGCGGCGATGGCCGCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGT 2212
Db      14431 TCGCCGCCCACTTCGAAGCGCGCGGGCGCAAGGCGACCCGCCTCCGCGTGAGCCACGCGT 14490

Qy       2213 TCCACTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGA 2272
Db      14491 TCCACTCACACCACATGGACCCCATGCTCGACGCGTTCCGCCGCGTCGCAGAAGGGCTCA 14550

Qy       2273 GCTACCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACG 2332
Db      14551 CGTTCCATCCGCCGCGCATCCCCATCGTCTCCAACGTCACGGGGAGGCTCGCGGCGGACG 14610

Qy       2333 AGGTGAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATG 2392
Db      14611 AGATCCGGTCGCCCGACTACTGGGTCCGCCACGTGCGAAGCGCCGTCCGCTTCGCCGACG 14670

Qy       2393 GAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGA 2452
Db      14671 GCATCAATACGCTGGAAGCCGACGGCGTTTCTTCGTTCTTCGAGCTCGGGCCGCACGGCG 14730

Qy       2453 CGCTGCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGC------CGGCGCTGCTCG 2506
Db      14731 TCCTCTCGGCGCTCGGCGAGAGCGCGCTGGCGAACGCGCAGCGCGAGGTCGCCTTCGTCC 14790
```

```
Qy        2507 CATCGTCGCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCT 2566

Db       14791 CCGCGCTCCGCGACGGCCGCGCCGACGTCGACGCGCTCACGGCGGCTCTCTCGTCGCTCC 14850

Qy        2567 GGGCCGTCGGTGGCCTGGTCTCCTGGGCCGGCCTCTT---CCCCTCAGGGGGGCGGCGGG 2623

Db       14851 ACGTGCGCGGACATCGCGTCGATTGGGCGGCGTTCTTCGGCCCGTTCGACCCCCAGACGG 14910

Qy        2624 TGCCGCTGCCCACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCGACG 2683

Db       14911 TCGCGCTCCCGACCTACGCGTTCCAGCGCGACCGCTTCTGGCTCGACGC---CCCCGCGG 14967

Qy        2684 ACGCGGCGCGTGGCGACCGCCGTGCTCCGGGAGCGGGTCACGACGAGGTCGAGGAGGGGG 2743

Db       14968 GAGCGACGCGCGACGAAGCCGCGCCCGCCGCGCCCGAAGACGCCGCGTTCTGGCGCGCGG 15027

Qy        2744 GCGCGGTGCGCGGCGGCGACCGGCGCAGCGCTCGGCTCGACCATCCGCCGCCCGAGAGCG 2803

Db       15028 TTGACGCGGGGGACGTGGGCGCGCTCGGCGCGACGCTGAACGCCTCGGGCGAGGACCACC 15087

Qy        2804 GACGCCGGGAGAAGGTCGAGGCCGCCGGCGACCGTCCGTTCCGGCTCGAGATCGATGAGC 2863

Db       15088 TCTCGGCCCTCGCGACGCTGCTCCCCGCGCTGTCGGCGTGGCGGCGCGCCCGCGACGAGC 15147

Qy        2864 CAGGCGTGCTTGATCACCTCGTGCTTCGGGTCACGGAGCGGCGCGCCCCTGGTCTGGGCG 2923

Db       15148 GGAGCCTCGTCGACGCGCTCCGCTACCGCGTCGTGTGGAAGCCGCTCACTCCGACCGCGG 15207

Qy        2924 AGGTCGAGATCGCCGTCGACGCGGCGGGGCTCAGCTTCAATGATGTCCAGCTCGCGCTGG 2983

Db       15208 CGCGCGACGTCGCGGGGACGTGGCTCCTCGTCACGCTCGCGGGAGCCGAGGACGCGCTCG 15267

Qy        2984 GCATGGTGCCCGA 2996

Db       15268 CGCGTGAGCTCGA 15280


RESULT 13
AF319998/c
LOCUS       AF319998                49736 bp    DNA     linear   BCT 28-MAY-2001
DEFINITION  Stigmatella aurantiaca myxalamid biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF319998
VERSION     AF319998.1  GI:14210834
KEYWORDS    .
SOURCE      Stigmatella aurantiaca
  ORGANISM  Stigmatella aurantiaca
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE   1  (bases 1 to 49736)
  AUTHORS   Silakowski,B., Nordsiek,G., Kunze,B., Blocker,H. and Muller,R.
  TITLE     Novel features in a combined polyketide synthase/non-ribosomal
            peptide synthetase: the myxalamid biosynthetic gene cluster of the
            myxobacterium Stigmatella aurantiaca Sga15
  JOURNAL   Chem. Biol. 8 (1), 59-69 (2001)
  MEDLINE   21110452
   PUBMED   11182319
REFERENCE   2  (bases 1 to 49736)
  AUTHORS   Silakowski,B., Nordsiek,G., Bloecker,H. and Mueller,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2000) MX, GBF, Mascheroderweg 1, Braunschweig
            38124, Germany
FEATURES             Location/Qualifiers
     source          1..49736
                     /organism="Stigmatella aurantiaca"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:41"
     CDS             complement(2..1306)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative aspartokinase"
                     /protein_id="AAK57182.1"
                     /db_xref="GI:14210835"
                     /translation="MRPIGSARQAGRGTSALALIVQKYGGTSVGDTERIKNVAKRCIA
                     AQKAGHDVVVVVSAMSGETNRLLKLVSQITERPNEREQDVVATGEQVSIGLVALAIQ
                     TQKRKAVSFLGHQVRITTDSTFSKARIKSIDAERIVDALKKKNIVVVAGFQGQDEHGN
                     VTTLGRGGSDTTAVALAAALKADACEIYTDVDGVYTTDPNMVPAARKLERIAYEEMLE
                     LASVGAKVLQIRSVEFAMKYKVPLWVKSSFTDDPGTLVCEEDAAMEDVLVSGIAYDKN
                     EAKIAIRGVPDVPGVAAKIFGALDEQSIVVDLIVQNVSKDGRTDVTFTVGKADLTKAK
                     DVVKKIAKSVKAEGVETDDQVSKVSIVGVGMRNHSGVAAKMFTVLAGAGVNVQMISTS
                     EIKVSCVIHSNYTELAVRSLHTAFGLDKPEAGAVSEVVSLKG"
     CDS             1363..2895
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /product="putative histidine ammonium lyase"
                     /protein_id="AAK57183.1"
                     /db_xref="GI:14210836"
                     /translation="MSRPRLNIDGDTLKLEEILQVARHTVTVELAPAAAARVKAARDL
                     VDRVAAGDTPSYGINTGFGTLAEVRIDKKDLRELQRNLILSHAAGVGSPLPLPEARVL
                     LLLRCNVLAKGYSGIRPETLALALEMLNRDVVPVVPERGSVGASGDLAPLAHLALVFI
                     GEGEAFYKGERLPAAQALERAGLKPVVLEAKEGLALVNGTQAMCAVGTLLQLRAEMLA
                     DLADLAGAMTLEGLLGSHKPFIPEIQDVRAHEGQKACAAHLRELLADSALVESHVNCS
                     KVQDPYSLRCMPQVHGAAREGLSFARRILEVEINSATDNPLVFVETERIVSGGNFHGQ
                     PVSLALDVAAMALTQLSAISERRVEQLVNPALSGLPPFLAKNSGLNSGFMIAQVTSAA
                     LVAESRVLSHPASVDSIPSSAGREDHVSMGMTAALKGRQVADFTRSCLAIELLVAAQA
                     LDYRQPTRAGKGPQAAYELIRSKIPTMEKDRELHRDIAAVSALIDSGELLNAVRTATR
                     GQ"
     gene            complement(3275..7822)
                     /gene="mxaA"
     CDS             complement(3275..7822)
                     /gene="mxaA"
                     /note="non-ribosomal peptide synthetase; contains a
                     condensing, adenylation, and PCP reductase domains"
                     /codon_start=1
                     /transl_table=11
                     /product="MxaA"
                     /protein_id="AAK57184.1"
                     /db_xref="GI:14210837"
                     /translation="MSVRELLTTLAAKGVRLRLEGDKLQVEASKAVLTPELRAELSAN
                     KADILAFLRKHSSREETGSELGQPLARPEVIPLSSGQERLWFLDRLSPGTSQYNVHLG
                     LRARGALDTKALRRSLDELVRRHEVLRTHFPEVEGSPRQVIVSPDTGVALTIVELPGL
                     TESQREEEIYRRSDELAQQPFDLAKGPLLRVTVLALGPDDFALFVTKHHIITDGWSLG
                     VFVRELSMLYASFVRGQAAALPPVSMQFADSALRERAWLAGESGARERAYWKGKLKGL
                     PPLQLPVDHAVSTSTSHRGATVPLALSPALSDALRELASREGCSFFTVLFAAFSALLH
                     RYSGQVDFGVGTVIANRGSVPADLIGFIANTLALRCDLSGEPTFSQWLARAHKVVLEA
                     LDYQALPFSEVVQTVGASRDGGLNPLVRACFTLESIPAPTLDLPGTSWSFLNGAPDGS
                     VEGVAKFELSLILAASEKGLAGMLEYSREVFDASTVERMVGHFQVLLESIVAHPEVPL
                     SKLPLLTAEERGRLLNDWNGPVLDVPALCMHELVQAQAERTPQAVAVVSGQKTLTYAE
                     LNRRANQLAHHLRRLGIQKEERVGLCVERTEDIVIGLLAILKAGGAYVPLDPAYPKER
                     LALILEDAQVPVLLTQQRLVPELPATQARVVCLDKDWPTIGAESDTNPERITAPEAIA
                     YLIYTSGSTGKPKGVMIEHRNAVAFLIWAMSVFSPKELAGTLASTSICFDLSVFEIFT
                     PLCCGAKVIVAKNALELPELPAAREVTLINTVPSAMGALLRSGGVPSSVAIVNLAGEA
                     LAGALVDQIYQLDHVRDVFNLYGPSETTTYSTFTRVNRGQTPTIGRPVGNTQVYALDS
                     NREPMPIGVPGEVYIGGMGVARGYLGRPELTAERFVRSPFGGGSEARLYRTGDLARWL
                     PDGQLEYLGRMDHQVKLRGFRIELGEIGAVLMEHSGIRDAVVVVREGLGADKQLVAYV
                     VGRGEKAPEPAELRDYLKSKLPEYMVPFLFVGLDALPLTPNGKVDRAALPAPERTHSG
                     PAKEHVAPRTPGEESLAAIWRQVLGVEQIGAHDNFFELGGHSLLLYRVLVLARSASGA
                     DIPLRALLQAPTLEEMARAVEAAKTGSLPAHDVTVEMEADAVLDAEIALGKALPPVTG
                     ALRTILLTGATGFLGAFLLEELCRRTDARIYCLVRSKTEQEGMNRIRKNLESYSLWNE
                     ALAPRIVPVRGDIGQPLLGLSEKEFQRLSEEIDAIYHNGALVNFLYPYESMRAANVLG
                     TREILRLATRTRIKPLHYVSTVSVLPLGRKAPIREDEPLEGPSSLVGGYAQSKWVAEK
                     LVREASRRGLPVTILRPGRVTGHSRTGAWNTDDLVCRTLKGCVRMGVAPSVDALLDLT
                     PVDYVSSAIVDLSMRPESIGQTYHLVNPQFVRADEMWNYMRAFGYGLRVLPYDQWLSE
                     LGSAASSDSELGDLLMFLQQVPPEDRSVGGPRMVVCDSGDTLKALGGTGTSCPSVDAS
                     LISTYLSSLVHRGFLKAPEVR"
     gene            complement(7819..11520)
                     /gene="mxaB1"
     CDS             complement(7819..11520)
                     /gene="mxaB1"
                     /note="PKS; ketosynthase; acyltransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="MxaB1"
```

```
                /protein_id="AAK57185.1"
                /db_xref="GI:14210838"
                /translation="MRAGAGGHPLLGVSFTTSVQKENRFWESALSAKEPSYLADHRVG
                GAVVVPGAAYLEMALSAARELSGEKAYELVDTSFKEGLLLPEGQERVVQMALSGEGGD
                ALSFQLSSQKASTEPNAPAGWMAHVVGQIRPIEGGPSTQTPESFEAIRERCSEVIAQA
                AHYEELASRGVAYGPAFQGVQQIWRGRAEALGHVRLPEPLAARASAYRIHPALLDACF
                QLVAAAVPQEDLTPDAGPAVPVALAHLRLHEQPGTEVFCHVRLRPSEGARRGVYEFDL
                VLRDASGRILVEALGLRVQQLDAPAAEHAGKDLFFALEWRHASAVQQVPAEGPGASGR
                WLLIADGSELADTVESQLRERHGEVIRVDLAGAGQGRRVVDATSPQAFDGVLSEAFGA
                GSAPRGVIHLTGLEPREVDPIMGCGGVLHLVQALSRVSLASPPRLWLVTRGVHHGSGS
                RSASDVVQASLWGLGRTLAHEHSELKCTRVDVSAAMGTGDAGAALMRELLVADDEEEI
                SLRADGRHVGRIVRGSPGRAPSEAVVPAGGRPFHLERDAQGRLEFRAGARRPPGPGEL
                EIEVEVAALGGPGEGTGPERWEGAGEGIGCSGRVVAIGEGVKNVAVGEERIALVASGL
                GSHVSVPAGCTAIRPSSSSHGEAAALAATVLPAWYALFHLGRLEKRGRVWVLGATSGL
                GRAAIKLAKRAGAEVFAVATTDEQRASLRELGADQVMDPSDASFVSRVLDATGGRGVE
                LAIVAPGAVELERSLPTLSEGARVLDLRASGASARTGDANVAWCVVDVQDFARRRPER
                FASLWQEVKAAFEAGGLPSAAADAKALALKDPKARVVVPAADARRLRADGTYLVTGGL
                GGLGLAVAKWMVEQGAGHLVLVGRDTVLSPEQQEVVTALEAAGARVRVARADVSDRAQ
                IARILSEIAEGGAPLRGVIHAAGVLDDGVLLQQTVERFRRVMAPKVLGGWNLHVLTRE
                APLDFFVLYSSAASMFGAPGQGNYVAGNAFLDALAHHRRALGLPGLSINWGPFSEVGL
                AAAQANRGERLAQRGSDSLTPAEGNTILGRLLDGDVTQMAVMPLELRKWVEFYPRAKS
                SPWLSELVPADAGQEVGEARDVALLETLRTATPQEAKAALEQFVREQLARVLRLDSAR
                IDPEAPLQSFGLDSLMGLELRNRLVSGLGVSLPASLIWKHPTLAALCAHLLGEVMDRT
                LAETLAQAAAGAGPAVDESAADDKEVFVL"
gene            complement(11487. .14303)
                /gene="mxaB2"
CDS             complement(11487. .14303)
                /gene="mxaB2"
                /note="PKS; dehydratase; enoylreductase; ketoreductase;
                ACP"
                /codon_start=1
                /transl_table=11
                /product="MxaB2"
                /protein_id="AAK57186.1"
                /db_xref="GI:14210839"
                /translation="MSKQEKPAEQPSALQRAALLVEKMQARLDAVERARTEPIAVIGL
                GCRFPGGAVDGSSYWKILRDGVDALREVPESRWDVPGHFDSTRGVPGKMYGTRGGFLD
                DVEHFDAEFFGISPREAASLDPQQRLVLEVAWEALENAGLAPDQLAGSKTGVFMGVMS
                SDYMARLLKENDATRFDGYMATGNGYSFVPGRVSYVLGLQGPCMPVDTACSSSLVALH
                LASESLRAGESNLALAGGVNLILSPETMICLCSMQALASDGRCKTFDASADGYVRGEG
                CGVLVLKRLSDAQRDGDSILALIRGSAVNHDGASGGLTVPNGPSQQAVVQRALDNARI
                APALVGYIEAHGTGTPLGDPIELRALGAVLGKGRPADRPFFIGSVKTNIGHLEPAAGI
                AGVIKTILSLQHKEIPPHLHFRTPNPHVEWDRIPARVPVERVPWPAHEGRRIAGVSSF
                GLSGINAHVVLEEAPASTAPTAVAEEDKAQLVVLSAKTPRALSSLAEAWSSFLKNEEG
                GSLADISYTSALSRTPHEHRLALAAKSKQELIEQLEAFAAGKSHAGLSSGRRGTGLGQ
                KIVFVFPGQGSQWLGMGQRLYAENVTFREAIDWCHEAISRYTDWSLRDLLTNPGQQPR
                WNDIDVIQPTLFALQVALAAVWRWLGLEPHAVVGHSMGEVAAAHVAGALSLEDAARII
                CERSKLLRRVSGQGAMAVVDLSREQAEVEIRGAEDRIAIAVSNGPRSTVLSGDPAALK
                EVLERLERREVFCRWVKVDVASHSPQMDPLRQDLLDRMAAIRPTGSAVPLYSTVTGAP
                IDGRELSAPYWVRNLRQPVLFADAVQRLIQDGNVLFIELSPHPILVPFVDAMLREGDT
                SGRGLVVPSLIRDQEEWPSLLSSVGALFTRVIAWTGSGYIPTSVAGLRFLRTRGSASG
                TGWSCQQLRRRAGSCALAREVIPS"
gene            complement(14300. .29878)
                /gene="mxaC"
CDS             complement(14300. .29878)
                /gene="mxaC"
                /note="PKS; modules 1,2 and 3 contain KS, AT, DH, KR, and
                ACP regions"
                /codon_start=1
                /transl_table=11
                /product="MxaC"
                /protein_id="AAK57187.1"
                /db_xref="GI:14210840"
                /translation="MANRNDESNELSPAKRMLVALEKMQTRLNAVEGAAKEPIAIVGM
                ACRFPQADNVEAFWRLLQGGVDAVREIPLERWPKDEVADIGDGAMRWGGFLDSVDSFD
                PGFFGISPREAVRMDPQQRLLLEVAWEALEDAGLDVDKLSGSRSGVFIGACNDDYHCM
                QVERPETGDAFSATGVAASVLSGRLSYLFNLQGPSLVVDTACSSSLVSLHLACQSLRN
                RECNMALAGGVNLILSPQSVLLVSKLQALSPDGRSKAFDASANGFTRGEGCGIVVLKR
                LSDALADGDHILATIRGSAINQDGKSTGLTTPNVLSQQALIRQALESSGLKAEQVSYV
                EAHGTGTPLGDPIEAEALRETYGVPRPDGSVCGIGSVKTNVGHLESAAGIAGIMKVVL
                AMRHQSIPPHLHLKQVNPRIQLEGSALTIPTRLTPWKASGQPRRAGVSSFGISGTNAH
                VIIEEAPAPVAKDALPQRPEVLPLSARSREALVALAERYAERLGREGGGSLSDVCYTA
                SAHRTHHTHRLAVLADSREHTAERLRAFARGEQPSGAWSGRKAFGGRRKIVFVYPGQG
                AQRAGMGRALMESEPVFRDALVKVDKALQPHLGWSVLDEIAADEQRSRLAEIDVNQPV
                LFAVEVALTALWRSWGIEPDAVMGHSMGEVVAAHVAGALSLEDAVLIICHRSKLMRTL
                GGRGATMAMVELSVEEAEEAIRPWNGKIWLGGLNGPRSQVLSGEPEAVKEAVEALGKR
                GVFTRWVKMDVPSHTPLVEVISRELEGRIARITPRAADIPLCSTVTGEFLDGAKLDAA
                HWGNNLRNPVRFAAAAQHLLASDHGTFVEVSPHPVLLPAVEQTMSAGNTEGVVLASLR
                RNEDERHTMLSSLGALYSSGHRVAWNRVYPGKGQLVPLPTYPWQRERFWMEPAVRGVR
                ASRTSSGHPLIGERIDLSAQGEVSRLWQVQLGPKHFPLLAEHRVQGVPALPAAAMADM
                ALSAAKQTLGEPVSLEDVSFSALLSFPEDEERVLQLWLSTERPGASAFRLFSAPVMRE
                TPSDELTWALHAEGSLRHGDASVVEKAAPVALDELRARCGEEASVEAHYQTLKEIGLE
                YGPSFRVVRELYRAKGEALGRVQLEGELTAEEERYSLHPALLDGCFQVLAAALEGLDS
                EGGVYVPVALSRLRVLVRPEAALWSHVRIVSGAEAGSNTIECDVTLLNDQGRVVAEVR
                RLVCKKIAAASAPADEVDRWIYSAQWEPQQRQARAETGSSRTAGRWLVFADARGIGQS
                LASGLAERAGGCVVVEAGPSYERLEADRYRINPSKLEDFQALLKEAFSDQSPCKGMVH
                LYGLDVASNVTEPESFERALSLSSTSVVYLLQAWAQRGWRDTPRLWLVTTGVQVIASG
                EQPGSPAAAPLVGLAAVITHEHPELRCAHVDLSAQPSPEEIASLTEELLGDGAEDRIA
                LRGAVRHVARLGRGISGRKAAEPREFTADGTYLITGGLGGLGLEVAKWFIQRGARNLA
                LTGRGAPSEEAQKVLGELRANGAEVRVYRADVTSAADVARLLGSIDSELPPLRGVVHA
                AGVIADGVLTQLDEPRLRAVLAPKVQGAWNLHVQTQGRPLELFVMFSSAASMMGSPGQ
                GNYAAANAFMDVLAHHRRAAGLPGLSVNWGPWADVGLAAVATNRGDRLALRGVGNMPP
                AQALEALGRLLVQEAAQVAVVPLDLRQWREFYLSAAQSPFLSVLMQEKVSTPASRKGS
                ARELLAAAEPEQRKGILEGYLREQLGRILRMEPSRIEPDQPFGDLGMDSLTGLELRNR

  Query Match             19.8%;  Score 844.4;  DB 1;  Length 49736;
  Best Local Similarity   58.4%;  Pred. No. 2.9e-83;
  Matches 1563;  Conservative    0;  Mismatches 1091;  Indels   24;  Gaps    4;

Qy          1 GTGGCGGATCGTCCCATCGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGT 60
Db      19792 GAGGAGGCGAGCCGCGTCGTGGGCAACGCCACGGAGCCGATTGCGATCGTCGGCATTGGC 19733

Qy         61 TGCCGTCTGCCCGGTGGCGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCG 120
Db      19732 TGCCGAGTCCCGCAGGCTTCG---GGTCCGGAGGCGTTCTGGAAGCTGTTGGAGGGAGGC 19676

Qy        121 CGCGACACCGTCGGGCGAGTCCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCC 180
Db      19675 GTGGACGCCATCCGCGAAGTGCCTGCCTCGCGGTGGAAGGTCGAGGACTACTACGACCCG 19616

Qy        181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCC 240
Db      19615 CAGCCCGGCGTCCCGGGGAAGACCTACACCCGGTGGGGTGGTTTCTTGGATGGGGCTGAC 19556

Qy        241 TGCTTCGACGCCTCCTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCA 300
Db      19555 CAGTTCGACGCCCAGTTCTTCGGGATCGCCCCTCGCGAAGCGGCGAACATGGACCCGCAG 19496

Qy        301 CATCGACTCTTGCTGGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCG 360
Db      19495 CAGCGCCTCCTGCTGGAGGTGGCCTGGGAGTCCCTGGAGCACGCGGAATCGCTCCGGCG 19436

Qy        361 GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420
Db      19435 CGCCTGGCCAACACCCGCACTGGCGTGTTCATTGGTATTGGCAGCAACGAGTACGCGATG 19376

Qy        421 GCGCTGCCGCAAGCGACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGACGATG 480
Db      19375 CTCAACGGGGTGGGGAGCGCCTCGGCCGCAGGGGATGCCTACATCGCCACGGGCAATGAC 19316

Qy        481 CCCAGCGTCGGAGCGGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCG 540
Db      19315 TCGAGCTTCGCCGCGGGCCGGCTGGCCTATCTGCTCCGGTTCCAGGGGCCGACGATGAGC 19256

Qy        541 GTGGATACGGCCTATTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGC 600
Db      19255 TTGAACACCGCGTGCTCCTCTTCGCTGGTGGCCGTGCACCTGGCCTGCCAGAGCTTGCGC 19196

Qy        601 TCCGGGGAATGCTCCACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACC 660
Db      19195 GCCGGTGAGAGCAACCTGGCGCTGGCGGGTGGCGTCAACATGACGCTCGCGCCGCACTCG 19136

Qy        661 CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCG 720
Db      19135 ACCATCTACCTCGCGCAGCTTCGTGCCTTGTCTCAGGATGGGCGGTGCAAGACCTTCGAT 19076
```

```
Qy      721 GCGGAGGCCGATGGGTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTC 780
Db    19075 GCCTCCGCGGACGGCTACGTGCGCAGCGAGGGTTGCGGCATCGTGGTTCTCAAGCGCCTG 19016

Qy      781 AGTGGAGCCCGCGCGGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAAT 840
Db    19015 TCCGATGCGAAACGGGATGGGGATGACGTCCTGGCCGTCATCCGCGGCTCGGCGGTCAAC 18956

Qy      841 CACGACGGTGCGAGCAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTG 900
Db    18955 CACGATGGTCCGAGCAGTGCGCTCACGGTGCCCAATGGCGATGCCCAGCAGCAGGTTCTC 18896

Qy      901 AAACGGGCCCTGGCGGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACAC 960
Db    18895 CGGGCGGCCCTGGCCAACGCGGGGGTTGCGCCTGCGGACGTGGATTACATCGAGGCGCAT 18836

Qy      961 GGCACGGGCACGACGCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGC 1020
Db    18835 GGAACGGGCACGTCACTCGGCGATCCCATCGAGGTCCAGGCTCTGGCCAACGTGCTCGGC 18776

Qy     1021 CTCGGGCGAGATGTCGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCAT 1080
Db    18775 CAGGGACGTACGTCCGAACAGCGGCTGCTGATCGGCTCGGTGAAGACCACCATTGGTCAC 18716

Qy     1081 CCTGAGTATGCGTCGGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGG 1140
Db    18715 CTGGAGGCAGCGGCGGGCATCGCCGGTCTCATCAAGGTCGTCCTCGCCCATCGGCATGGG 18656

Qy     1141 CAGATTCCTGCGCACCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTT 1200
Db    18655 GTGGTTCCGCCGCACCTGCACCTCAAGAAGCTCAACCCGCACATCGAGCTGAATGGCTTC 18596

Qy     1201 CGGCTGACCGTCACGCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCG 1260
Db    18595 CCGCTGGACATCCCGACACAGGTCACTCCGTGGCCGGAGCGTGCGCGTCCTCGGATCGCT 18536

Qy     1261 GGGGTGAGCTCGTTCGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCG 1320
Db    18535 GGCATCAGCGCCTTTGGCTTGAGCGGCACCAATGCGCACGTCCTCGTGGAGGAGGCGCCC 18476

Qy     1321 GCGGCGACGTGCACACCGCCGGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCA 1380
Db    18475 GCTCCCGAGCCTCGCGCGGAACAGGCGGAGCGGAGCCTGCACGTGCTTGCGCTCTCGGCG 18416

Qy     1381 AGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTAC 1440
Db    18415 AAGACGAAGGGTGCGCTCACGCAACTCGCGGCTCGGGTGGGTGAGTACCTCGCCGAGCAG 18356

Qy     1441 CCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAG 1500
Db    18355 CCAGAGCAGAAGCTCGCGGATGTCTGCCATGCGGCGAACACGCAGCGCACTTGGTTCTCG 18296

Qy     1501 CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCG 1560
Db    18295 CATCGACTCGCGCTCGTCGTGGAGTCGGTGGAGCAGACCCGGGAGCGGCTGTTGGCGTTC 18236

Qy     1561 GCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG 1620
Db    18235 GCGCGAGGGGAGGAAGTCCCTCGGCTGGCCCAGAACCGCGTCGAAGGCGAAGATCCCAAG 18176

Qy     1621 CTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTAC 1680
Db    18175 GTCGTCTTCGTATTCACGGGGCAGGGCGCGCAGTACGCGGGCATGGGCCGGGAGCTGTAC 18116

Qy     1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAG 1740
Db    18115 GAGACCGAGCCTGTCTTCCGTGCGGCCTTGGATCGGTGCAATGCATTGATCGAGGGCAAG 18056

Qy     1741 CTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTG 1800
Db    18055 TGGGCCGAGTCCCTCCTCTCGGTGATGC---------------ACGGCACGGGCTCGCGG 18011

Qy     1801 CTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCG 1860
Db    18010 ATTGACGACACCGAGTACACCCAGCCGGCCTTGTTCGCGTTGGAGTACGCCCTGGCGGAG 17951

Qy     1861 CTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTG 1920
Db    17950 CTGTGGCGTTCCTGGGGCGTCGAGCCCTGGGCGGTGTTGGGCCACAGCGTCGGTGAGTAC 17891

Qy     1921 GTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCG 1980
Db    17890 GTGGCCGCCTGCGTGGCGGGGGTGATGGAGCTCGAAGAAGCCCTCGGTCTCATCACCGAG 17831

Qy     1981 CGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCG 2040
Db    17830 CGTGCGCGGCTCATGCAAGCACTGCCGCGCAATGGCGAGATGGTGGCCGTCTCCGCGAGT 17771

Qy     2041 GAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTC 2100
Db    17770 GAGGCGCAGGTCCAGGAAATGCTGGCTCCTTATGCGGATCGGGTTTCGATCGCGGCGGTC 17711

Qy     2101 AACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCG 2160
Db    17710 AATGGCCCGGCGGACACCGTGATTTCCGGTGAGAGGGAGGCGGTGGAAGCCATCTCCGCT 17651

Qy     2161 GCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCA 2220
Db    17650 TCGTTGCAGGCCAAGGGCGTCAAGGCCCGGCGGCTTACCGTGTCCCACGCGTTCCACTCT 17591

Qy     2221 CCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGG 2280
Db    17590 CCGCTCATGGACCCCATGCTGGATGCGTTCGAGAAGGCGGCTGGCAAGATCGCCTTCCGT 17531

Qy     2281 CGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGA---CGAGGTG 2337
Db    17530 GCGCCTCGCCTCCACCTGGCCTCCAACGTCACGGGCCAGCTCCTCGGAGATGGCGCCGTG 17471

Qy     2338 AGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTG 2397
Db    17470 CTCGACGCCGCATACTTCCGCCGCCACGTGCGCGAGGCCGTGCGCTTCCACGACGGGCTG 17411

Qy     2398 AAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTG 2457
Db    17410 AAGGCACTTCGGGCGCTCGGTGGCACCGTGTTCGTGGAGGTCGGCCCGCACCCGACGCTC 17351

Qy     2458 CTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGC 2517
Db    17350 ACCAGCATGGCGCTGAAGGGGCTGGGGGATGATGCGGGCACTTGGTTACCCTCGCTGCGC 17291

Qy     2518 GCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGT 2577
Db    17290 AAGGGTCAGCCCGAGCAGCCGCAGATTCTTTCCAGCCTGGGCGCACTCTACACCCAGGGC 17231

Qy     2578 GGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGG---GGGGCGGCGGGTGCCGCTGCCC 2634
Db    17230 GTGGCGGTGAACTGGATCGCGCTGAACGGCGAGCGCCACCGGTCGCCCGTTGCGCTCCCT 17171

Qy     2635 ACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACAC 2672
Db    17170 GTCTACCCGTGGCAGCACCAGCGGCACTGGCTGGAAAC 17133
```

RESULT 14
SAU421825

```
LOCUS       SAU421825            66808 bp    DNA     linear   BCT 16-MAR-2002
DEFINITION  Stigmatella aurantiaca ORF8, ORF7, ORF6, ORF5, ORF4, ORF3, ORF2,
            ORF1, stiA gene, stiB gene, stiC gene, stiD gene, stiE gene, stiF
            gene, stiG gene, stiH gene, stiJ gene, stiK gene, stiL gene and
            ORF9.
ACCESSION   AJ421825
VERSION     AJ421825.1  GI:19572309
KEYWORDS    acyl-CoA binding protein; cellulase; cytochrome P450 monooxygenase;
            kinase; methyl transferase; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6;
            ORF7; ORF8; ORF9; polynucleotide adenyltransferase; potassium
```

```
            channel beta chain; ribosome binding factor; stiA gene; StiA
            protein; stiB gene; StiB protein; stiC gene; StiC protein; stiD
            gene; StiD protein; stiE gene; StiE protein; stiF gene; StiF
            protein; stiG gene; StiG protein; stiH gene; StiH protein; stiJ
            gene; StiJ protein; stiK gene; stiL gene.
SOURCE      Stigmatella aurantiaca
  ORGANISM  Stigmatella aurantiaca
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE   1
  AUTHORS   Gaitatzis,N., Silakowski,B., Kunze,B., Nordsiek,G., Blocker,H.,
            Hofle,G. and Muller,R.
  TITLE     The biosynthesis of the aromatic myxobacterial electron transport
            inhibitor stigmatellin is directed by a novel type of modular
            polyketide synthase
  JOURNAL   Online Publication
  REMARK    J. Biol. Chem., 10.1074/jbc.M111738200
REFERENCE   2  (bases 1 to 66808)
  AUTHORS   Muller,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124
            Braunschweig, GERMANY
FEATURES             Location/Qualifiers
     source          1. .66808
                     /organism="Stigmatella aurantiaca"
                     /mol_type="genomic DNA"
                     /strain="Sg a15"
                     /db_xref="taxon:41"
     CDS             complement(4978. .6231)
                     /note="ORF8"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P450 monooxygenase"
                     /protein_id="CAD19077.1"
                     /db_xref="GI:19572310"
                     /translation="MLRSVSASRPPSPANPVAAFDLARLDDAFYADPFPLYRAMRERD
                     PVHRMPDGSLFLTRWADLDCIYRDTRTFSSDKRAEFGAKYGDAPLFEHHTTSLVFNDP
                     PLHTRVRRLIVGALTPRALSTMEPGLRTLVDRLLDGLAVKGAADLIEDFAAAIPIEVI
                     GNLLDIPTEERGPLRGWSLAILGALEPRLTAEQEACGNEAVTEFLDYLRILVAQRRAR
                     PGDPATDVLTRLIQGESDGERLTETELLHQCVFLLNAGHETTTNLIGNALELLARFPD
                     ERARLLRAPALIPTAVEEVLRYESSNQLGNRRVAEDTEIGGVAVPTGTFLTLCIGAAN
                     RDPARFEDPEHFDVGRQPNRHLAFAGGAHTCAGMNLARIEARIALAAFLARFPDYALT
                     APPVRARRARFRGFTAMPTRLGSLR"
     CDS             6369. .7730
                     /note="ORF7"
                     /codon_start=1
                     /transl_table=11
                     /product="protein kinase"
                     /protein_id="CAD19078.1"
                     /db_xref="GI:19572311"
                     /translation="MEELISEALPDGSRVGPWRIEGRAGYGTYGAVYRAHRVGRICPR
                     SVALKLARYPNDLRFEREAELLERVRHANVPRLLGHGTWRAGPQGEAHPYVVMQWVEG
                     VRLYQWNKERLRTAREVLRVLAQVARALEATHMRGGLHRDVKGENVLVSPEGHASLMD
                     FGCGTWEGAAPLTGGILAPGTKAYRSPEALRFHWNHRNEADAHYQATPADDVYALGVT
                     AYRLCTGIYPPPATEPSGVGEDERETQEVKKPPSQVRPLVPALEQLILRMLRDNPKER
                     GRAGELAAEMEALAERAGAEADEPMNPNHSSVPSEITGRPVSVQREWGGWPVALVLVA
                     VGLLLWGSGSLNTGRRVAPASRMTEGGTGGVADAAVEALPVSSAAREPEPSGLNLDML
                     KEPLPGQRRPPCPRFQISIQGGCWLEVPQSSPPCGESYYEWKGACYSPVAAPPRPRTS
                     DPQ"
     CDS             complement(7849. .8577)
                     /note="ORF6"
                     /codon_start=1
                     /transl_table=11
                     /product="methyl transferase"
                     /protein_id="CAD19079.1"
                     /db_xref="GI:19572312"
                     /translation="MSQNIYDDPSFFESYSQLARSLKGLAGAPEWPALRALLPDVKDL
                     RVVDLGCGFGWFCRWARKNDAAHVLGLDLSEKMLTRAKAETSGDAITYQQADLERLVL
                     PEGAFDLAFSSLAIHYVEELEPLLATVHRGLTPGGWFVFSIEHPIFMASRQPEWMTGP
                     EGHRTWPVDSYQLEGKRTTHWLAEEVVKYHRTLGTTLNALLEQGFTLRHIEEWGPTDA
                     QLSAHPELAEERDRPMFLLVSAQR"
     CDS             complement(8622. .8879)
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="acyl-CoA binding protein"
                     /protein_id="CAD19080.1"
                     /db_xref="GI:19572313"
                     /translation="MALEDDFRSAQERVKALKTRPSNDRLLELYSLFKQATEGDVQGK
                     RPGILDVKERAKYDAWVGRKGLGRESAMQQYVALVERLLRG"
     CDS             8961. .10070
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="potassium channel beta chain"
                     /protein_id="CAD19081.1"
                     /db_xref="GI:19572314"
                     /translation="MVYLLCRSRSQLRWILNGSAHRVRFAWGGHARPRRMHFRHLGRS
                     GLMVSEISYGNWLTHGSQVEEAAAVACVRAALDEGITTFDTADVYAGTRAEEVLGRAL
                     QGQRREGLEIPTKVYWPTGPGPNDRGLSRKHILESIHGSLRRLKTEYVDLYQAHRYDT
                     ETPLEETMQAFSDVVRQGKAHYIGVSEWTADQLREGVKLSRQMGFQLVSSQPQYSMLW
                     RVIEAEVVPASQELGVGQVVYSPIAQGVLTGKYRPGQKPPEGSRATDEKGGASFIKRF
                     LNDDVLTRVQQLTPVAQEAGLSLAQLAVAWVLQNPNVSSAIIGASRPEQVKENVKAAG
                     VKLSEEAMRHIDTLLGPIVERNPKKTLENSPRKRP"
     CDS             10428. .10862
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /product="ribosome binding factor"
                     /protein_id="CAD19082.1"
                     /db_xref="GI:19572315"
                     /translation="MTSFKQRRRRAPRRGTESSLSSFEAQSLPAHHLRVQSTLFQEVA
                     RLFRGELSDPLLEGVSVTSFELTPDGRNARIGFTLPPEYAEGGSAPIEAALVRASGFI
                     RAQLALGLNLKRVPHLRFVCVGVAPPGFSIPEDGSAEEGGES"
     CDS             10859. .12016
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD19083.1"
                     /db_xref="GI:19572316"
                     /translation="MRPRRIEGAPGAVPILAWARAVPAGAVKQLQLLARQPYLAEHVA
                     AMPDVHVAEGVAVGTVFATERHIVPGALGGDLGCGVSAVRFDFPAMSLGRVELTRLLA
                     AFARAVPVGDAVHRGRGLSLPPGLDIGRLSTHRLQREWERLVPKHLGTLGGGNHFLEL
                     DRDAEGALWLLLHTGSRGVGAAIAAHHLRVAQVSGEGSLPGLRTDTSEGAACLADTEL
                     ACQFARANRDTLLQRAVEVLAGAIGQGPVLDSRVDVHHNHVAAESHFGRTLLVHRKGA
                     VGLVAGQQGLIPGSMGTASYLVAGKGEPQAFCSCSHGAGRVMTRREARERIRPAALAH
                     ALRRVVFDEGRVTALVEEAPTAYRDITEVLEDEAELVTPLRRLTPIAVLKG"
     CDS             complement(12082. .13974)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="cellulase"
                     /protein_id="CAD19084.1"
                     /db_xref="GI:19572317"
                     /translation="MKNSLLLLMLAVLPLYGTAALARPGMANAEPGHFEGPENAGLTA
                     AQCNPRVPLSTRGRYIVDRCGERFKLKSVNWFGASDQLEVVGGLDKQKLSDLVAGMKA
                     MGFNSVRLPFSNNMLHPDDNKPAAQRCLEKYSVTCIDSTKNPELLNKTALEVYDAVVA
                     ELTRQGLVVILNNHTTKSMWCCGWDGNGFWDSSQALQRWQDDWVMLAARYQGNKWVAG
                     ADLRNEVRPDGLDSPNWGMRNQHDWHMAAQTMGNLLLRTNPDLLIVVEAVNWWGLLDG
                     SRPQLKPVRQRPVALLRGDKLVYAVHNYGYTGPNQSGGSLGSGPKYSDMDKPTLYGTL
                     DQEWGFALAANQAYTAPVWMSEFGVGYNEQAANSRAWFSNLADYLIEKDVDWAYWAIN
                     AAKLQNSVQGEDETYGLWSYPDWNGVRSGDWRLGTGPLGRLLGADGRTGAVDATRFLP
                     MTVLLKDGSSTYYVDNNSLDSDWHSGKAKISCPRGYRMVGVSANFTILCTNAGAVPAQ
                     QGTGSYQTITQEVSPLRYPGDWANQSIKFECPTGFYAVGISTANPFFLELAGLLCEQN
                     TGGIPLNSKSAKNFWGGDQRASSSGGDWSVSQYKGQCADNQYLIGLAHRRSGLADYPS
                     VVLCSN"
     gene            14345. .21466
                     /gene="stiA"
     CDS             14345. .21466
                     /gene="stiA"
                     /function="PKS"
                     /codon_start=1
                     /transl_table=11
                     /product="StiA protein"
                     /protein_id="CAD19085.1"
                     /db_xref="GI:19572318"
```

```
                    /translation="MTQLASRLGVGSESIDPRERFSRYGIDSRDTKSLLAELGALLKR
                    PLSPTLAWTYPTVEALSRHLAGEADVSGGERADTLGVDQDEPIAIIGMAGRFPKAPDV
                    ESFWSLLRDGVDAITEVPPERWDLDAFYSVGLKSPGKMNTRWGGFLDQVDQFDPLFFG
                    ISPREAIQMDPQQRLALELSWEALESAGVAVDSLMDSRTGVFMGAMWSDYAKLTRGAL
                    GLIGPHTATGQDSSIIAARVSYFLGLRGPSLTVNTACSSALVAVHLACQSLRNGEAAM
                    ALAGGVNLLLSPESSVAMSQLGALSPDGRCKAFDARANGYVRGEGGVVVLTPLSRAI
                    ANGYPVLCVIRGSAVNNDGFSNGLTAPNPRAQEEVLRDAYARAGVSPDKVHYVETHGT
                    GTVLGDPIEANALSAVLGVNRPAERPLILGSVKTNIGHLEAASGIAGLMKVVLSIQHR
                    AIPPNLHFEQPNPYIRFEELRLLVPKSLHPWPNEGESALAGISSFGFGGTNCHLVLEA
                    PSPGSAHIAPLAAETPEELSLLARQARDAVSVVGSQVELADLCFSVAARCREGSQRLA
                    LIARSPAELAAQVDGFLAGEKRPGLLRGEVTSGGSQKPVFVFSGTGSQWPGMGRELLR
                    TEPVFRGAIERCDRALRGLSGWSMLEELSAGEASSRLKECSGLQTAIFAIQVGLAALW
                    RSWGIEPAGVVGHSIGEVAAAHVAGILNLEDAARVIHHLSRLMGRLSGQGAMIVVDLS
                    AEEANRLLREWERHVVVAACNAPASTVLSGAREVIEELGAELERRGIAAHTVRIDTPS
                    HSPLMDPFMGELAEVLAGLSPRVAAVPMVSTVTGAFLKAEECTAAYWVRNLREPVLFS
                    QALTKLFEQDLRLFLELSPHPILARSVEQCLKGEGRQGVVLASMRRGENEGACMRESL
                    GRLFTLGTSVAWQALYPAEAKRVELPMGLGTGSLDAELRGSARLLPLSAHTAEALEDA
                    ARALQASIAKHPDVRLDDLCYSASVGKSVHAHRLAFVTDSKRALVEHLDDFLKGESRP
                    GRIAGVCATGEPKKVVFVFPGQGSQWQGMGRQLLRQEPVFREAIEECDAAIRSCAGWS
                    LIDALQADEPHSRLGSIEVIQPTLFAIEVALAALWRSWGVVPDAVVGHSMGETAAACV
                    AGALRLTDAARVICERSQLLRRVAGQGGMALVELSLEQAREMLSRKDGRVSIGASNSP
                    RSTVLSGDAIALEEILRELEAQGTFCRRVKVDVASHSPQVDPLKDDLLRVLSTIAPRA
                    ATVPLYSTVTGEPIEGARMDASYWMRNLREPVLFGQAIERLAGDGHGVFVELSPHPIL
                    LPSIDQTLRHKSHKGVVLPSLRREEPERAVLLESLGTLFTGGGRVDFRGLHPLGGHRV
                    RLPHLPWQRQRFWLDMDVAAAGAASVGAAWHPGGGSEHPLLGHRLPSLAHAPACHVWE
                    NKLSRSAVSYLDDHRVGGTVVLPGTSYIEMALAAAAKGSGKKTVSLRDVELREKMSFL
                    ETASRTVQLSLIQEEAGHSSFRLHGRSTPDEPWTLHGGTVHVEEGDGPPSADPVLWA
                    AVETVRLQGKEVSASDAYEAMHAQGNEYGPLFRGIEQMWVRDAELVARLRVPEPLEQE
                    LPEYRFHPAALDLGLQALIVLAAEHGCEGLALPTRLESIRAHASPGARMWCHVRLNPV
                    REAASFSGEVLFLDESGQRLAELQGARFQILGEGSRPHDGGRLVDWLHEVEWQPAPVP
                    VPTAEPSSQGRWVLFADGTGVAEGLAAQIAARGEPCIRVLPGMAFAEAEDGHFVVDPE
                    QPEDVRKLFGALSRTSTIPLRGIVHLWSLDAPSPEEGGVDALSRARRLGSGSVLSLIQ
                    GLNAAELAQKPRLWLVTRGVQAVGSAGPGAGAVAQSPLWGLGQVVAYEHAELRCTRVD
                    LAPVSVPGEEQTLLQELLAPAGDDQVALRGRERHVPRLTRCRTDEVAGRSTDLRAGAI
                    RSDGCYLITGGLGGLGLTIAKWLVARGARHLVLVGRRGASDAAEIALGQMRGAGAECL
                    VMKADVANTSEVEELVGTIQRKMPPLRGIFHAAGLLDDGTVLQLTQARLRSVMAPKVE
                    GAWNLHRATLGASLDFFVLFSSAASLVGIPGQGNYAAGNAFLDALAGLRKARGLACQS
                    IQWGPWSEIGLAAEQDNRGERLALRGMGSLNPREGLEALEALLSPGEPLRERARIGVL
                    RLDARRWCKSYPAIRSSPLFSVLLAGERNAKQVLERSGLRARLNEEPARTRRAALESF
                    LREEAAQVLRLPPARIDLSRSLQTLGMDSLLSLELRNRLEANLELTLPATVLWNYPTL
                    SALAPFLAERMGIQLNGVEEPPQAVESPSVSEEDGAAMLKMLEQFQELSNEDLLGMLS
                    DDKEDEVSS"
     gene           21463..26220
                    /gene="stiB"

  Query Match             19.2%;  Score 817.2;  DB 1;  Length 66808;
  Best Local Similarity   57.5%;  Pred. No. 2.4e-80;
  Matches 1597;  Conservative    0;  Mismatches 1148;  Indels   31;  Gaps    6;

Qy         21 GCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGGCGT 80
Db      43614 GCGTCTGGCCCCCGAGCCCATCGCGATCGTGGGCATGGCCTGTCGCTACCCGGTGGGGT 43673

Qy         81 GATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGAGT 140
Db      43674 GAAGGACTCCGACTCCTATTGGGATTTGTTGAAGCGCGGCGGAGATGCCGTCACGGAGAT 43733

Qy        141 CCCCGCCGAACGCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCCGGGGAA 200
Db      43734 TCCAGCAGCCCGCTGGGACATCAACGCCTTCTACGATCCGAACCAGGACAGCCCCGGCAA 43793

Qy        201 GACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTCCTTCTT 260
Db      43794 GACGACGACGCGCTGGGGGGCTTCCTGGATGATGTGGAGGGGTTCGATCCCCACCTGTT 43853

Qy        261 CGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCTGGAGGT 320
Db      43854 CGGCATCTCGCCTCGGGAAGCGGCGAGCATGGATCCCCACCAGCGCCTCTTCCTGGAGGT 43913

Qy        321 GTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTACGGAAAC 380
Db      43914 CGCCTGGGAGGCGCTCGAGAACGCAGGCCAGCAAACCGATCGTCTGGCCGGGAGCAAGAC 43973

Qy        381 GGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGCGACGGC 440
Db      43974 GGGGGTCTTCGTCGGCCTTTACACGAGCGACTACTCTCCTGTTGTTC------TCTCGGA 44027

Qy        441 GTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGCGGGCCG 500
Db      44028 TCTCAACGCCATCAACGTCTATACCGGCACAGGCACGGCCAACAACGTGGCGGCTGGACG 44087

Qy        501 AATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTATTCGTC 560
Db      44088 ATTGTCCTATCTGCTGGACCTGCAAGGGCCCAGCTTTGTGGTGGACACGGCCTGCTCCTC 44147

Qy        561 CTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTCCACGGC 620
Db      44148 CTCGCTTGTCGCCGTTCACCTGGCGTGTCAGAGCCTTCTGAATCGTGAATGCAGAATGGC 44207

Qy        621 CCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTCGAAGAC 680
Db      44208 GGTCGTGGGCGGCGTGACGTTGTTGCTCACGCCACTGACCCTCGTCATGGCGTCGAAGAT 44267

Qy        681 CCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGGGTTCGG 740
Db      44268 GCAGCTCATGGCCGCGGACGGCCGCTGCAAGACGTTCGACAGCCGCGCCGATGGTATCGC 44327

Qy        741 ACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGCGGACGG 800
Db      44328 CATGGGTGAGGGCTGTGGCGTGGTGATCCTCAAGCGCTTGTCGGATGCCCAGGCCGATGG 44387

Qy        801 CGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAGCAGCGG 860
Db      44388 TGATCGCATCCTGGCGGTGATCAAGGCTTCGGCGATCAATCAAGACGGCCGCACGAATGG 44447

Qy        861 TCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGCCCTGGCGGACGC 920
Db      44448 GCTGACAGCGCCGAGTGGCGCTTCCCAGCAGGTGTTGTTGCGTCAGGCGCTGACGCGCGC 44507

Qy        921 AGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGACGCTTGG 980
Db      44508 TGGCATCTCGCCGTCGCAGGTGACGTACGTGGAAGCCCACGGTACGGGGACACCGCTCGG 44567

Qy        981 TGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGTCGCCAC 1040
Db      44568 CGATCCGATTGAGGTGGAGGCCCTCGCGGAGGTCATCGGGCAGCCACGACCTTCCGGGCC 44627

Qy       1041 GCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGAT 1100
Db      44628 TCCCTGTGTCCTGGGTTCCGTGAAGGCCAATCTGGGACATACCGGCGCGGCGGCGGGGGT 44687

Qy       1101 CACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCACCTCCA 1160
Db      44688 CGCGGGGCTGATCAAAGTCGTGCTCTCCATGCAGAACGAGGCGATTCCTCCTCAACTGCA 44747

Qy       1161 CGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCACGCGCGC 1220
Db      44748 CTACCGGGAGCTCAACTCCAACATCTCCTTGGAGAACACCTCCCTGGTCATTCCGAAGGA 44807

Qy       1221 CCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTTCGGCAT 1280
Db      44808 GCTGCATCCCTGGCCTGCGGGGAAGGGGCCGCGTATCGCAGGGATCAGTGCCTTCGGTTG 44867

Qy       1281 GAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCACACCGCC 1340
Db      44868 GTCAGGGACCAACGCGCACGTGCTTGTGGAGGAGGCGCCTGCGGCCCCCAAGGCTCGCGC 44927

Qy       1341 GGCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTCGGCAAGGACCGCGTCAGCCCTGGA 1400
Db      44928 GGCGGTTGAGCGTCCGGTGCACCTCCTGACCCTGTCGGCTCGGACCGAGGCCGCGCTGCG 44987

Qy       1401 TGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGACCTACCCTTCGCAGTGTCTGGGCGA 1460
Db      44988 CGAGTTCGCGGAGCGTGTGAAGACATATGTCTCGGCGCACGCGGATCTGCCCATCGCCGA 45047

Qy       1461 TGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGATGGAGCACCGGCTCGCGGTGGCGGC 1520
```

```
Db     45048 TCTCTGCTACTCGGCGAACACGGGGTGGTCGAAGTTCGAGCACCATCTGGCGGTCCCCGC 45107

Qy      1521 GACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGCTGCGGCGCAGGGACAGACGTCGCC 1580

Db     45108 GGGGACGACGGCCCAGTTGAGCGAGCGTCTCTCCGCGTTTCTTGAGGGACAGCCGCAAGC 45167

Qy      1581 CGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGGCAAG---CTCGCCTTTCTCTTCAC 1637

Db     45168 AGGCACGTCGTACGGACAGGTCACGGGACGCCGTGGGCCGGGAGTCGCGTTCCTGTTCAC 45227

Qy      1638 CGGACAGGGGGCGCAGACGCTGGGCATGGGCCGTGGGCTGTACGATGTATGGTCCGCGTT 1697

Db     45228 GGGGCAGGGCTCGCAGTACGCGGGCATGGGGCGGCAACTCTACCAGACACAGCCCACATT 45287

Qy      1698 CCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCAGGAGCTCGACCGGCCGCTCCG 1757

Db     45288 CCGGCGTGCCCTCGATCGGTGCGACGAGTTGCTGCGCCCCTTGCTCGGGCGCTCCATTCT 45347

Qy      1758 CGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGCGCTGCTCGACCAGACAGCCTT 1817

Db     45348 GTCGGTGATCTACCCCGAAGCG------GGGCAGGCTTCTCCGATCGACGAGACGGCCTT 45401

Qy      1818 CACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGCCGCGCTGTGGCGGTCGTGGGG 1877

Db     45402 CACGCAGCCCGCGTTGTTCGCGTTCGAGTGGGCGCTCGCGGAACTCTGGCGCTCGTGGGG 45461

Qy      1878 TGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGAGCTGGTGGCTGCCTGCGTGGC 1937

Db     45462 CGTTGTACCGACCGTGGTGATGGGCCACAGCGTGGGTGAGTATGTGGCGGCCTGCGTCGC 45521

Qy      1938 GGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGCTGCGCGCGGGCGCCTGATGCA 1997

Db     45522 TGGGGTGTTCACCCTCGAAGAAGGGCTTGCGCTCATCGTGGAGCGCGCCCGGTTGATGCA 45581

Qy      1998 GGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGCGCCGGAGGCCGATGTGGCTGC 2057

Db     45582 GTCGTTGCCGCCGGGCGGCGAGATGGCCGCGGTCTTCACGGACGGGGCTCGTACCGCGCA 45641

Qy      2058 TGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGCGGTCAACGCTCCGGACCAGGT 2117

Db     45642 GGCGATTGCTCCTTACGCATCGCAGGTCTCCATTGCCGCCTTCAATGGTCCTTCCGAGAC 45701

Qy      2118 GGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGCGGCGGCGATGGCCGCGCGCGG 2177

Db     45702 GGTGATCTCTGGCGATGGCAAAGCCGTGGAGGCGATTCTCGGGGCGTTGTCCGCCGAGGG 45761

Qy      2178 GGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCACTCACCGCTCATGGCCCCGAT 2237

Db     45762 CGTGAAGGTTCGGCGCCTGGGTGTCTCCCACGCCTTCCACTCTCCGCTCATGGACCCCAT 45821

Qy      2238 GCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTACCGGCGGCCGTCGATCGTCCT 2297

Db     45822 GCTGGATGCGTTCGAGCGGGCGGCGGCGAAGACGCGCTTCATGCCCCCGAAGATCACGCT 45881

Qy      2298 GGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGTGAGCTCGCCGGGCTATTGGGT 2357

Db     45882 GATCTCGAACCTCACCGG------TGGGCCCTCGGAGGAGTTCTCGGCCCGGGCGTTGCG 45935

Qy      2358 GCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGTGAAGGCGCTGCACGCGGCCGG 2417

Db     45936 TCGACATGCGCGTGAACCGGTCCGGTTCTCGGATGGGATGGCCGCGCTCAAGGCACGTGG 45995

Qy      2418 TGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCTGCTCGGCCTGGTGCCTGCCTG 2477

Db     45996 CGTCTCGGTCGCGGTGGAGATTGGCCCGAATGCAACTCTGCTCGGCATCGGAAGCCGGTG 46055

Qy      2478 CATGCCGGACGCCCGGCCGGCGCTGCTCGCATCGTCGCGCGCTGGGCGTGACGAGCCGGC 2537

Db     46056 TCTGCCGGAAGGGTCGGTGGCTTGGTTGCCTTCCCTTCGAAAGGACAAAGACGAGTGGGA 46115

Qy      2538 GACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGTCGGTGGCCTGGTCTCCTGGGCCGG 2597

Db     46116 AGTCCTGCTCGCAACGTTGGGTGCGCTCTCTGTGCGAGGCGTTCCCGTCAACTGGGCAGG 46175

Qy      2598 ---CCTCTTCCCCTCAGGGGGGCGGCGGGTGCCGCTGCCCACGTACCCTTGGCAGCGCGA 2654

Db     46176 GTTCGACAGCGACTACCCGCGTCGGCGGGTCTCCCTGCCCAATTATCCGTTTCAGCACGA 46235

Qy      2655 GCGCTACTGGATCGACACGAAAGCCGACGACGCGGCGCGTGGCGACCGCCGTGCTCCGGG 2714

Db     46236 GCGGTTCGCGCTC-------AAGCAGATGGATGGACGGGGTGTGGCGCCGAGCGGTCAGG 46288

Qy      2715 AGCGGGTCACGACGAGGTCGAGGAGGGGGGCGCGGTGCGCGGCGGCGACCGGCGCAGCGC 2774

Db     46289 TCGTGGGTCATCCCCTGTTGGGATCGCGCCTGCGCGTCGCGACGGTGGACGGCACTCTGT 46348

Qy      2775 TCGGCTCGACCATCCG 2790

Db     46349 TCGAATCGAAGTTCCG 46364
```

RESULT 15
AR266876
LOCUS AR266876 3978 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 19 from patent US 6495348.
ACCESSION AR266876
VERSION AR266876.1 GI:29696258
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 3978)
AUTHORS Sherman,D.H., Mao,Y., Varoglu,M., He,M. and Sheldon,P.
TITLE Mitomycin biosynthetic gene cluster
JOURNAL Patent: US 6495348-A 19 17-DEC-2002;
FEATURES Location/Qualifiers
source 1..3978
/organism="unknown"
BASE COUNT 497 a 1583 c 1415 g 483 t
ORIGIN

Query Match 19.1%; Score 814.2; DB 6; Length 3978;
Best Local Similarity 57.6%; Pred. No. 1.2e-79;
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

```
Qy        18 CGAGCGCGCAGCCGAAGATCCGATTGCGATCGTCGGAGCGAGTTGCCGTCTGCCCGGTGG 77

Db        90 CGAGGACGCGGCCCGCGAGCCCGTCGCGATCGTGGGCATGGCGTGCCGCTTCCCCGGGGA 149

Qy        78 CGTGATCGATCTGAGCGGGTTCTGGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCG 137

Db       150 CGTGGCATCGCCGGACGACCTGTGGCAGCTGGTCGCCGAGGGCCGGGACGCCGTCACCGA 209

Qy       138 AGTCCCCGCCGAAC---GCTGGGATGCAGCAGCGTGGTTTGATCCCGACCCCGATGCCCC 194

Db       210 GTTCCCCGCCGACCGGGGCTGGGACGTCGACGCCGTCTACGACCCCGAGCCGGGCACCCC 269

Qy       195 GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTGAGCGACGTAGCCTGCTTCGACGCCTC 254

Db       270 GGGCAGGACGTACGCGCGCCACGGCGGCTTCCTCAAGGACGCCGCCGGATTCGACGCCGC 329

Qy       255 CTTCTTCGGCATCTCGCCTCGCGAAGCGCTGCGGATGGACCCTGCACATCGACTCTTGCT 314

Db       330 CTTCTTCGGCATCACGCCGCGCGAGGCGCTCGCCATGGACCCGCAGCAGCGCATGATCAT 389

Qy       315 GGAGGTGTGCTGGGAGGCGCTGGAGAACGCCGCGATCGCTCCATCGGCGCTCGTCGGTAC 374

Db       390 GGAGGTCTCCTGGGAGGCGTTCGAGCAGGCGGGCCTCGACGCGACCACCCTGCGGGGCGA 449

Qy       375 GGAAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCCGCGCTGCCGCAAGC 434

Db       450 GGACGTCGGCGTCTTCGTCGGCTCCAACAGCAACGACTACCTGATCAACGTGCTCGACGC 509

Qy       435 GACGGCGTCCGCAGAGATCGACGCTCATGGCGGGCTGGGGACGATGCCCAGCGTCGGAGC 494
```

```
Db    510 GCGGGACGTCGCCGAGGGCTTC------ATCGGGACCGGCAACTCCGCCAGCATCCTCTC 563

Qy    495 GGGCCGAATCTCGTATGCCCTCGGGCTGCGAGGGCCGTGTGTCGCGGTGGATACGGCCTA 554
Db    564 CGGCCGCGTCGCCTACACCTTCGGCTTCGAGGGCCCGGCCGTGTCCGTCGACACCGCCTG 623

Qy    555 TTCGTCCTCGCTGGTGGCCGTTCATCTGGCCTGTCAGAGCTTGCGCTCCGGGGAATGCTC 614
Db    624 CTCCTCCTCGCTGGTCGCGCTGCACCTGGCCGCGCAGTCCCTGCGGCAGGGGGAGTGCTC 683

Qy    615 CACGGCCCTGGCTGGTGGGGTATCGCTGATGTTGTCGCCGAGCACCCTCGTGTGGCTCTC 674
Db    684 CCTGGCGCTGGCGGGCGGCGCGACGGTGATGGCCACGCCGACCGCCTTCATCGAGTTCAG 743

Qy    675 GAAGACCCGGGCGCTGGCCAGGGACGGTCGCTGCAAGGCATTTTCGGCGGAGGCCGATGG 734
Db    744 CCGCCAGCGGGGCCTGGCCCCCGACGGCCGCTGCAAGTCCTTCTCGGCGACCGCCGACGG 803

Qy    735 GTTCGGACGAGGCGAAGGGTGCGCCGTCGTGGTCCTCAAGCGGCTCAGTGGAGCCCGCGC 794
Db    804 CACCACCTGGTCCGAGGGCGCGGCCGTGCTGCTGCTGGCCCGGCTCTCGGACGCCCGCCG 863

Qy    795 GGACGGCGATCGGATATTGGCGGTGATTCGAGGATCCGCGATCAATCACGACGGTGCGAG 854
Db    864 CCTGGGCTACCCCGTGCACGCGGTCATCCGGGGCAGCGCCGTCAACCAGGACGGCGCGAG 923

Qy    855 CAGCGGTCTGACCGTGCCGAACGGGAGCTCCCAAGAAATCGTGCTGAAACGGGCCCTGGC 914
Db    924 CGCGGGCCTGACCGCGCCCAACGGACCGGCGCAACAGCGGGTGATCCGGCAGGCACTGGC 983

Qy    915 GGACGCAGGCTGCGCCGCGTCTTCGGTGGGTTATGTCGAGGCACACGGCACGGGCACGAC 974
Db    984 CAACGCACGGCTGACGGCCGACAGCGTCGACGCGGTCGAGGCACACGGCACCGGCACCCC 1043

Qy    975 GCTTGGTGACCCCATCGAAATCCAAGCTCTGAATGCGGTATACGGCCTCGGGCGAGATGT 1034
Db   1044 GCTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGGCGGGCCCGCGGCGA 1103

Qy   1035 CGCCACGCCGCTGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTC 1094
Db   1104 GGGCAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCCACACCCAGTCCGCGGC 1163

Qy   1095 GGGGATCACTGGGCTGCTGAAGGTCGTCTTGTCCCTTCAGCACGGGCAGATTCCTGCGCA 1154
Db   1164 CGGCGCGGGCGGCGTCATCAAGATGGTGATGGCCATGCGGCACGGGACGCTGCCCCGCAC 1223

Qy   1155 CCTCCACGCGCAGGCGCTGAACCCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTCAC 1214
Db   1224 GCTGCACCTCACGGAGCCCACCCCGCGCGTCGACTGGTCCGCCGGTGACGTACGGCTGCT 1283

Qy   1215 GCGCGCCCGGACACCGTGGCCGGACTGGAATACGCCGCGACGGGCGGGGGTGAGCTCGTT 1274
Db   1284 GACCGAGGCCCAGGACTGGCCGGACACCGGACAGCCGCGCCGTGCGGCCGTCTCGTCCTT 1343

Qy   1275 CGGCATGAGCGGGACCAACGCGCACGTGGTGCTGGAAGAGGCGCCGGCGGCGACGTGCAC 1334
Db   1344 CGGCGTCAGCGGCACCAACGCCCATGTGATCCTGGAGGGCCCGCCCGCCGAGGAGGCACC 1403

Qy   1335 ACCGCCG------------------GCGCCGGAGCGACCGGCAGAGCTGCTGGTGCTGTC 1376
Db   1404 GGACGCGCCGCTGCCGGACGTCTCCTCGCAGCCGCGGGGCCCGCTGCCGTGGGTCGTCTC 1463

Qy   1377 GGCAAGGACCGCGTCAGCCCTGGATGCACAGGCGGCGCGGCTGCGCGACCATCTGGAGAC 1436
Db   1464 CGGCCGCAGCGAGGCGGCCGTCCGAGCGCAGGCCGAGCGCCTGGCGGCCCACCTGACCGC 1523

Qy   1437 CTACCCTTCGCAGTGTCTGGGCGATGTGGCGTTCAGTCTGGCGACGACGCGCAGCGCGAT 1496
Db   1524 GCGCCCGCACCTGGCACCGGCCGACGTGGCCACCGCGCTGGCCACCACGCGGGCGGCCTT 1583

Qy   1497 GGAGCACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGCGGGCAGCCCTGGACGC 1556
Db   1584 CGACCACCGGGCCGCCGTCGTCGGCCGGGACCGTGAGGAACTGCTCGCCGGCCTCGCGGC 1643

Qy   1557 TGCGGCGCAGGGACAGACGTCGCCCGGTGCGGTGCGCAGTATCGCCGATTCCTCACGCGG 1616
Db   1644 CCTGGCCACCGGAACCCGCGCGCCCGGCCTGGTCACCGGCCGGACCCCGCCGTCCGGCGG 1703

Qy   1617 CAAGCTCGCCTTTCTCTTCACCGGACAGGGGGCGCAGACGCTGGGCATGGCCGTGGGCT 1676
Db   1704 CAAGGCCGCCTTCCTCTTCACCGGACAGGGCAGCCAGCAGCCCGGCATGGGCCGCGAACT 1763

Qy   1677 GTACGATGTATGGTCCGCGTTCCGCGAGGCGTTCGACCTGTGCGTGAGGCTGTTCAACCA 1736
Db   1764 GGCGGCTCACAGCACCGTGTTCGCCGACGCCCTGGACGAGGTCTGCGCCCAGCTCGACCG 1823

Qy   1737 GGAGCTCGACCGGCCGCTCCGCGAGGTGATGTGGGCCGAACCGGCCAGCGTCGACGCCGC 1796
Db   1824 GCACCTCGACCGGCCGCTGCGCGAGGTGCTGTTCGCCGCGGACGGCACGCCCGAGGCCGC 1883

Qy   1797 GCTGCTCGACCAGACAGCCTTCACCCAGCCGGCGCTGTTCACCTTCGAATATGCGCTCGC 1856
Db   1884 CCTGCTCGACACGACGGCCTACACCCAGCCCGCGCTGTTCGCCGTCGAGGTCGCGCTGCT 1943

Qy   1857 CGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAGTTGGTCGCCGGCCATAGCATCGGTGA 1916
Db   1944 GCGGCTGCTGGAGGACTGGGGCTTGCGGCCCGGCATGGTCGCGGGCCACTCGGTCGGCGA 2003

Qy   1917 GCTGGTGGCTGCCTGCGTGGCGGGCGTGTTCTCGCTTGAGGACGCGGTGTTCCTGGTGGC 1976
Db   2004 ACTGACCGCCGCCTACGCCGCCGGGGTCTGGTCGCTCGCCGACGCCTGCGCCCTGGTCGC 2063

Qy   1977 TGCGCGCGGGCGCCTGATGCAGGCGCTGCCGGCCGGCGGGGCGATGGTGTCGATCGAGGC 2036
Db   2064 CGCCCGCGGCCGGCTGACCCAGGCACTGCCCGCGGGCGGCGCCATGGTCGCCGTGCAGGC 2123

Qy   2037 GCCGGAGGCCGATGTGGCTGCTGCGGTGGCGCCGCACGCAGCGTCGGTGTCGATCGCCGC 2096
Db   2124 GACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCCGC 2183

Qy   2097 GGTCAACGCTCCGGACCAGGTGGTCATCGCGGGCGCCGGGCAACCCGTGCATGCGATCGC 2156
Db   2184 CGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCGACGAGGCCGCCGTCACGGACCTGGC 2243

Qy   2157 GGCGGCGATGGCCGCGCGCGGGGCGCGAACCAAGGCGCTCCACGTCTCGCATGCGTTCCA 2216
Db   2244 GCGCGAGTGGGCCGCCCGCGGCCGGGAGACCAGGAGGCTGCGGGTCAGCCACGCCTTCCA 2303

Qy   2217 CTCACCGCTCATGGCCCCGATGCTGGAGGCGTTCGGGCGTGTGGCCGAGTCGGTGAGCTA 2276
Db   2304 CTCCGCCCACCTGGACGCCATGACCGAGGCGTTCGCCGAGGTCGCACGAGGGGTGTCCTA 2363

Qy   2277 CCGGCGGCCGTCGATCGTCCTGGTCAGCAATCTGAGCGGGAAGGCTTGCACAGACGAGGT 2336
Db   2364 CAGCGCGCCGTCCCTCCCGGTGGTCTCCACGCTCACCGGGGCCCCCGTCACCGACGAGCT 2423

Qy   2337 GAGCTCGCCGGGCTATTGGGTGCGCCACGCGCGAGAGGTGGTGCGCTTCGCGGATGGAGT 2396
Db   2424 CCGCAGGCCGGAACACTGGGTGCGGCACGTCCGGGAGACGGTGCGCTTCCACGACGCGGT 2483

Qy   2397 GAAGGCGCTGCACGCGGCCGGTGCGGGCACCTTCGTCGAGGTCGGTCCGAAATCGACGCT 2456
Db   2484 CCGCGCCCTGCGCGACCGCGGGGCCACCGCGTTCCTGGAGGTCGGGCCCGGCGGCGTGCT 2543

Qy   2457 GCTCGGCCTGGTGCCTGCCTGCATGCCGGACGCCCGGCC---GGCGCTGCTCGCATCGTC 2513
Db   2544 GACGGCCGCGGCACGCCGATGCCTGCCCGACGCCGCCCCCGAGACGTTCGTCCCCGTGCT 2603

Qy   2514 GCGCGCTGGGCGTGACGAGCCGGCGACCGTGCTCGAGGCGCTCGGCGGGCTCTGGGCCGT 2573
Db   2604 GCGGCGCCGCAGGCCCGAACCCGAGTCCGTGCTGACGGCCGTCGCGCAGGCCCACACGAT 2663

Qy   2574 CGGTGGCCTGGTCTCCTGGGCCGGCCTCTTCCCCTCAGGGGGGCGGCGGGTGCCCGCTGCC 2633
Db   2664 CGGCCTCTCGCCGGCGTGGGACCGCCTGCTGCCCAAGGCCCGGACGCGCGTGGACCTGCC 2723
```

```
Qy       2634 CACGTACCCTTGGCAGCGCGAGCGCTACTGGATCGACACGAAAGCCG 2680
              ||||||| | |  |||||||  | ||||||| | |     |  ||||
Db       2724 CACGTACGCCTTCCAGCGCGGCCACTACTGGCTGGCGGGCATGGCCG 2770
```

Search completed: October 4, 2003, 03:30:21
Job time : 15291.8 secs

GenCore version 5.1.6


```
OM protein - protein search, using sw model

Run on:         October  2, 2003, 17:27:29 ; Search time 125.882 Seconds
                                             (without alignments)
                                             2912.986 Million cell updates/sec

Title:          US-10-014-717-2
Perfect score:  7210
Sequence:       1 VADRPIERAAEDPIAIVGAS..........GVQNDFVSSGADQDWEIIAL 1421

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:      830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_23:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7207	100.0	1421	2	Q9L8C9	Q9l8c9 polyangium
2	7081	98.2	1421	2	Q9KJ00	Q9kj00 polyangium
3	4137	57.4	2439	2	Q9L8C5	Q9l8c5 polyangium
4	4098	56.8	2439	2	Q9KIZ5	Q9kiz5 polyangium
5	3322.5	46.1	7257	2	Q9KIZ7	Q9kiz7 polyangium
6	3311.5	45.9	7257	2	Q9L8C7	Q9l8c7 polyangium
7	2634	36.5	2218	2	Q8RJY1	Q8rjy1 stigmatella
8	2606	36.1	6315	2	Q9ADL6	Q9adl6 polyangium
9	2416.5	33.5	5017	2	Q8GBX6	Q8gbx6 polyangium
10	2357	32.7	5644	2	Q93NX8	Q93nx8 streptomyce
11	2355.5	32.7	5435	2	Q9L4X2	Q9l4x2 streptomyce
12	2331.5	32.3	11096	2	Q9L4W3	Q9l4w3 streptomyce
13	2330	32.3	9507	2	Q9EWA1	Q9ewa1 streptomyce
14	2328.5	32.3	10917	2	Q93NW6	Q93nw6 streptomyce
15	2321	32.2	4151	16	O53490	O53490 mycobacteri
16	2313	32.1	3513	2	Q8GBX5	Q8gbx5 polyangium
17	2311.5	32.1	6146	2	Q93HJ5	Q93hj5 streptomyce
18	2298.5	31.9	2103	16	Q9CD81	Q9cd81 mycobacteri
19	2275.5	31.6	1259	2	Q8RJX8	Q8rjx8 stigmatella
20	2273	31.5	2152	2	Q9ALM5	Q9alm5 saccharopol
21	2262.5	31.4	1885	2	Q8RJY4	Q8rjy4 stigmatella
22	2260.5	31.4	9477	2	Q9L4X3	Q9l4x3 streptomyce
23	2260	31.3	8817	2	Q53840	Q53840 polyangium
24	2258	31.3	9510	2	Q93NX9	Q93nx9 streptomyce
25	2257	31.3	1604	2	Q8RJX9	Q8rjx9 stigmatella
26	2255.5	31.3	1585	2	Q8RJY5	Q8rjy5 stigmatella
27	2249	31.2	6145	2	Q93H84	Q93h84 streptomyce
28	2245.5	31.1	3798	2	Q9L8C6	Q9l8c6 polyangium
29	2238.5	31.0	3798	2	Q9KIZ6	Q9kiz6 polyangium
30	2229.5	30.9	5192	2	Q93TW9	Q93tw9 stigmatella
31	2211.5	30.7	2478	16	Q8YTN5	Q8ytn5 anabaena sp
32	2207	30.6	1602	16	O65933	O65933 mycobacteri
33	2199.5	30.5	3352	2	Q93H83	Q93h83 streptomyce
34	2174	30.2	3579	2	Q8GBX4	Q8gbx4 polyangium
35	2165	30.0	6797	2	Q9X993	Q9x993 streptomyce
36	2157.5	29.9	1360	2	Q9RFK6	Q9rfk6 stigmatella
37	2157.5	29.9	3524	2	Q93H86	Q93h86 streptomyce
38	2156.5	29.9	6048	2	Q93H87	Q93h87 streptomyce
39	2152	29.8	1835	2	Q93H85	Q93h85 streptomyce
40	2150	29.8	1841	2	O33958	O33958 streptomyce
41	2136	29.6	2518	16	Q8YTN4	Q8ytn4 anabaena sp
42	2128	29.5	3816	2	Q9KIV3	Q9kiv3 streptomyce
43	2127	29.5	3170	2	Q9ALM4	Q9alm4 saccharopol
44	2123.5	29.5	2024	2	Q9EWA3	Q9ewa3 streptomyce
45	2120	29.4	1587	16	Q8YWG5	Q8ywg5 anabaena sp

ALIGNMENTS

```
RESULT 1
Q9L8C9
ID   Q9L8C9      PRELIMINARY;      PRT;  1421 AA.
AC   Q9L8C9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polyketide synthase.
GN   EPOA.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=So ce90;
RX   MEDLINE=20130945; PubMed=10662695;
RA   Molnar I., Schupp T., Ono M., Zirkle R.E., Milnamow M.,
RA   Nowak-Thompson B., Engel N., Toupet C., Stratmann A., Cyr D.D.,
RA   Gorlach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RT   "The biosynthetic gene cluster for the microtubule-stabilizing agents
RT   epothilones A and B from Sorangium cellulosum So ce90.";
RL   Chem. Biol. 7:97-109(2000).
DR   EMBL; AF210843; AAF26919.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW   Phosphopantetheine.
SQ   SEQUENCE   1421 AA;  149035 MW;  A8B5615EDAEDC996 CRC64;
```

```
Query Match             100.0%;  Score 7207;  DB 2;  Length 1421;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 1420;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60

Qy         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120

Qy        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180

Qy        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240

Qy        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Qy        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360

Qy        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Qy        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480

Qy        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660

Qy        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720

Qy        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780

Qy        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840

Qy        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900

Qy        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960

Qy        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020

Qy       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080

Qy       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140

Qy       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200

Qy       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260

Qy       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320

Qy       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380

Qy       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
              |||||||||||||||||||||||||||||||||||||||||
Db       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421


RESULT 2
Q9KJ00
ID   Q9KJ00      PRELIMINARY;      PRT;  1421 AA.
AC   Q9KJ00;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   EpoA.
GN   EPOA.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SMP44;
RX   MEDLINE=20293058; PubMed=10831849;
RA   Julien B., Shah S., Ziermann R., Goldman R., Katz L., Khosla C.;
RT   "Isolation and characterization of the epothilone biosynthetic gene
RT   cluster from Sorangium cellulosum.";
RL   Gene 249:153-160(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SMP44;
RX   MEDLINE=20115953; PubMed=10649995;
RA   Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C., Julien B.;
RT   "Cloning and heterologous expression of the epothilone gene cluster.";
RL   Science 287:640-642(2000).
DR   EMBL; AF217189; AAF62880.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW   Phosphopantetheine.
SQ   SEQUENCE   1421 AA;  148744 MW;  3D055DDBF686E2C5 CRC64;
```

```
  Query Match             98.2%;  Score 7081;  DB 2;  Length 1421;
  Best Local Similarity   98.2%;  Pred. No. 0;
  Matches 1396;  Conservative    7;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
              :|||||||||||||||||| |||||||||||||||||||||||||:||||||||||||||
Db          1 MADRPIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTVGQVPAERWDAAAWFDP 60

Qy         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120

Qy        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
              |||||||||||||||||||||||:|||||||||||||||||||||||||| |||||||||
Db        121 ALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTMPSVGAGRISYVLGLRGPCVA 180

Qy        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALATDGRCKAFS 240

Qy        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Qy        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360

Qy        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Qy        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480
              |||||||||||||||||||||||||||||||||||||||||||:|||| |||||||||||
Db        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTAAALDAHAARLRDHLETY 480

Qy        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
              ||||||||||||||||||||||||||||| ||||||||||||||| || || ||||||||
Db        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSSEGLRAALDAAAQGQTPPGVVRGIADSSRGK 540

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        541 LAFLFTGQGAQTLGMGRGLYDVWPAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660

Qy        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
              ||||||||||||||||| ||||||||||||||||||||||| ||||||||||||||||||
Db        661 RGRLMQALPAGGAMVSIAAPEADVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAA 720

Qy        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVS 780

Qy        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840
              ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRA 840

Qy        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900

Qy        901 RRAPGAHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
              ||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RRAPGAHDEVEKGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDR 960

Qy        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020

Qy       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYAL 1080

Qy       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DGIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140

Qy       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200

Qy       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1201 PFLRNLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260

Qy       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320

Qy       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLTQN 1380

Qy       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
              ||||||||||||||||||||||||:||||||||||||||||
Db       1381 LLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL 1421


RESULT 3
Q9L8C5
ID   Q9L8C5      PRELIMINARY;      PRT;  2439 AA.
AC   Q9L8C5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Polyketide synthase.
GN   EPOE.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=So ce90;
RX   MEDLINE=20130945; PubMed=10662695;
RA   Molnar I., Schupp T., Ono M., Zirkle R.E., Milnamow M.,
RA   Nowak-Thompson B., Engel N., Toupet C., Stratmann A., Cyr D.D.,
RA   Gorlach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RT   "The biosynthetic gene cluster for the microtubule-stabilizing agents
RT   epothilones A and B from Sorangium cellulosum So ce90.";
RL   Chem. Biol. 7:97-109(2000).
DR   EMBL; AF210843; AAF26923.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   2439 AA;  257732 MW;  08A32D75BCBADF2C CRC64;

  Query Match             57.4%;  Score 4137;  DB 2;  Length 2439;
```

```
Best Local Similarity   43.4%;  Pred. No. 3.9e-194;
Matches  933;  Conservative  135;  Mismatches  308;  Indels  776;  Gaps   13;

Qy      6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65

Db     26 LEQERTEPIAIVGIGCRFPGGADTPEAFWELLDSGRDAV--QPLDRRWALVGVHPSEEVP 83

Qy     66 GKTPVTRASFLSD-VACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124

Db     84 -----RWAGLLTEAVDGFDAAFFGTSPREARSLDPQQRLLLEVTWEGLEDAGIAPQSLDG 138

Qy    125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184

Db    139 SRTGVFLGACSSDYSHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPCLTVDTA 197

Qy    185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244

Db    198 CSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHCRTFDASAN 257

Qy    245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304

Db    258 GFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLREAL 317

Qy    305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364

Db    318 QSARVDAGAIGYVETHGTGTSLGDPIEVEALRAVLGPARADGSRCVLGAVKTNLGHLEGA 377

Qy    365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424

Db    378 AGVAGLIKAALALHHELIPRNLHFHTLNPRIRIEGTALALATEPVPWPRAGRPRFAGVSA 437

Qy    425 FGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQC 484

Db    438 FGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSAAALDAQAARLSAHIAAYPEQG 497

Qy    485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544

Db    498 LGDVAFSLVSTRSPMEHRLAVAATSREALRSALEVAAQGQTPAGAARGRAASSPGKLAFL 557

Qy    545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604

Db    558 FAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCEVMWAEPGSSRSSLLDQT 617

Qy    605 AFTQPALFTFEYALALALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664

Db    618 AFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAGVFSLEDAVRLVVARGRL 677

Qy    665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724

Db    678 MQALPAGGAMVSIAAPEADVAAAVAPHAALVSIAAVNGPEQVVIAGAEKFVQQIAAAFAA 737

Qy    725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784

Db    738 RGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSIALVSNLSGKPCTDEVSAPGY 797

Qy    785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844

Db    798 WVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACLPDARPVLLPASRAGRDE 857

Qy    845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI----DTKADDAAR-- 898

Db    858 AASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERYWIEAPVDREADGTGRAR 917

Qy    899 ------------------------------------------------------------ 898

Db    918 AGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEVVFPGAGYLEMALSSGAE 977

Qy    899 -------------------------------------------------GDRRAP----- 904

Db    978 ILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQVASREPGERRAPFRIHA 1037

Qy    905 ------------------------------------------------------------ 904

Db   1038 RGVLRRIGRVETPARSNLAALRARLHAAVPAAAIYGALAEMGLQYGPALRGLAELWRGEG 1097

Qy    905 ------------------------------------------------------------ 904

Db   1098 EALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSP 1157

Qy    905 ------------------------------------------------------------ 904

Db   1158 GELWCHARVVSDGQQASSRWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLE 1217

Qy    905 ------------------------------------GAGH-------------------- 908

Db   1218 LDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVVHAAGDDTSTAGMRALLA 1277

Qy    909 ------------------------------------------------------------ 908

Db   1278 NAFDGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDADALESALMRGCDSVLSL 1337

Qy    909 ------------------------------------------------------------ 908

Db   1338 VQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDP 1397

Qy    909 ----------------DEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEI 952

Db   1398 AEPEGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPDAQRREKVEPAGDRPFRLEI 1457

Qy    953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012

Db   1458 DEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALGVAPNDLPGEEIEPLVLG 1517

Qy   1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072

Db   1518 SECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLA 1577

Qy   1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132

Db   1578 YLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAEVYATADTPENRAYLESL 1637

Qy   1133 GVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY 1192

Db   1638 GVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLMVLRACGRLVKLGRRDDC 1697

Qy   1193 ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGV------------- 1239

Db   1698 ADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLVAAGAISPLGSGLRVGGS 1757

Qy   1240 FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHA------------- 1286

Db   1758 LTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRAPAESSVAVRADGTYLVT 1817

Qy   1287 -----------------GAG---------------------------------------- 1289

Db   1818 GGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALEAHGARVTVAKADVADRS 1877

Qy   1290 ------------------------------------------------------------ 1289

Db   1878 QIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARFRTVMGPKVQGALHLHTLTREA 1937

Qy   1290 ------------------------------------------------------------ 1289

Db   1938 PLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAQGLPALSIDWGMFTEVGMAVA 1997

Qy   1290 --------------------------------------------------PST------- 1292

Db   1998 QENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITPRQWVEFYPATAASRRLS 2057

Qy   1293 --------------GDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALF 1338

Db   2058 RLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQVSHVLRLPEDKIEVDAPL 2117

Qy   1339 TRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL-DALATAL 1389

Db   2118 SSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWLLDDALVVRL 2169
```

```
RESULT 4
Q9KIZ5
ID   Q9KIZ5      PRELIMINARY;      PRT;  2439 AA.
AC   Q9KIZ5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   EpoF.
GN   EPOF.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SMP44;
RX   MEDLINE=20293058; PubMed=10831849;
RA   Julien B., Shah S., Ziermann R., Goldman R., Katz L., Khosla C.;
RT   "Isolation and characterization of the epothilone biosynthetic gene
RT   cluster from Sorangium cellulosum.";
RL   Gene 249:153-160(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SMP44;
RX   MEDLINE=20115953; PubMed=10649995;
RA   Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C., Julien B.;
RT   "Cloning and heterologous expression of the epothilone gene cluster.";
RL   Science 287:640-642(2000).
DR   EMBL; AF217189; AAF62885.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   2439 AA;  257507 MW;  D3A972BC49FC3497 CRC64;

  Query Match             56.8%;  Score 4098;  DB 2;  Length 2439;
  Best Local Similarity   42.8%;  Pred. No. 3.2e-192;
  Matches  922;  Conservative  139;  Mismatches  315;  Indels  776;  Gaps   11;

Qy          6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65
Db         26 LEQERTEPIAIIGIGCRFPGGADTPEAFWELLDSGRDAV--QPLDRRWALVGVHPSEEVP 83

Qy         66 GKTPVTRASFLSD-VACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124
Db         84 -----RWAGLLTEAVDGFDAAFFGTSPREARSLDPQQRLLLEVTWEGLEDAGIAPQSLDG 138

Qy        125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184
Db        139 SRTGVFLGACSSDYSHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPCLTVDTA 197

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244
Db        198 CSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHCRTFDASAN 257

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304
Db        258 GFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLRQAL 317

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364
Db        318 QSARVDAGAIDYVETHGTGTSLGDPIEVDALRAVMGPARADGSRCVLGAVKTNLGHLEGA 377

Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424
Db        378 AGVAGLIKAALALHHESIPRNLHFHTLNPRIRIEGTALALATEPVPWPRAGRPRFAGVSA 437

Qy        425 FGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQC 484
Db        438 FGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSTAALDAQAARLSAHIAAYPEQG 497

Qy        485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544
Db        498 LGDVAFSLVATRSPMEHRLAVAATSREALRSALEAAAQGQTPAGAARGRAASSPGKLAFL 557

Qy        545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604
Db        558 FAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCEVMWAEPGSSRSSLLDQT 617

Qy        605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664
Db        618 AFTQPALFALEYALAALFRSWGVEPELIAGHSLGELVAACVAGVFSLEDAVRLVVARGRL 677

Qy        665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724
Db        678 MQALPAGGAMVSIAAPEADVAAAVAPHAASVSIAAVNGPEQVVIAGAEKFVQQIAAAFAA 737

Qy        725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784
Db        738 RGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSMALVSNLSGKPCTDEVCAPGY 797

Qy        785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844
Db        798 WVRHAREAVRFADGVKALHAAGAGIFVEVGPKPALLGLLPACLPDARPVLLPASRAGRDE 857

Qy        845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRA- 903
Db        858 AASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERYWIEAPVDGEADGIGRAQ 917

Qy        904 -----------------------------------------PGAGH---------- 908
Db        918 AGDHPLLGEAFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEVVFPGAGYLEMALSSGAE 977

Qy        909 ------------------------------------------------------------ 908
Db        978 ILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQVASREPGARRASFRIHA 1037

Qy        909 ------------------------------------------------------------ 908
Db       1038 RGVLRRVGRAETPARLNLAALRARLHAAVPAAAIYGALAEMGLQYGPALRGLAELWRGEG 1097

Qy        909 ------------------------------------------------------------ 908
Db       1098 EALGRVRLPESAGSATAYQLHPVLLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSP 1157

Qy        909 ------------------------------------------------------------ 908
Db       1158 GELWCHARVVSDGQQAPSRWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLE 1217

Qy        909 ------------------------------------------------------------ 908
Db       1218 LDWEPAALEGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVVHAAGDDTSAAGMRALLA 1277

Qy        909 ------------------------------------------------------------ 908
Db       1278 NAFDGQAPTAVVHLSSLDGGGQLDPGLGAQGALDAPRSPDVDADALESALMRGCDSVLSL 1337

Qy        909 ------------------------------------------------------------ 908
Db       1338 VQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDP 1397
```

```
Qy        909 ----------------DEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEI 952
                              |: || |:||:| ||| | ||: ||||: ||||||||||
Db       1398 AQPEGEADALLAELLADDAEEEVALRGGERFVARLVHRLPEAQRREKIAPAGDRPFRLEI 1457

Qy        953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012
              |||||| |||| | ||||| ||||||:|||| |:|||:|: |:|||| | :||
Db       1458 DEPGVLDQLVLRATGRRAPGPGEVEIAVEAAGLDSIDIQLAVGVAPNDLPGGEIEPSVLG 1517

Qy       1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072
               ||||||||||||||||||||||:|| |||||||| |||||| ||| ||||||:|
Db       1518 SECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLA 1577

Qy       1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132
              ||||||||::| || ||||| | ||:|| ||:||| |||||:||| ||||||||||
Db       1578 YLTAWYALDKVAHLQAGERVLIRAEAGGIGLCAVRWAQRVGAEVYATADTPEKRAYLESL 1637

Qy       1133 GVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY 1192
              ||||||||| || ||| ||| ||||||||:||||| |||| :||: || |:||:||
Db       1638 GVRYVSDSRSGRFAADVHAWTDGEGVDVVLDSLSGEHIDKSLMVLRACGRLVKLGRRDDC 1697

Qy       1193 ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGV------------- 1239
               || | || ||| ||| ||||||||::||||:||||:|| |:|||
Db       1698 ADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLVAAGAISPLGSGLRVGGS 1757

Qy       1240 FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHA------------- 1286
               ||||: | ||:| |:||| ||| ||||||||| ||||:|| |  :
Db       1758 LTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRAPAESSVAVRADGTYLVT 1817

Qy       1287 -----------------GAG---------------------------------------- 1289
                               |||
Db       1818 GGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALEAHGARVTVAKADVADRS 1877

Qy       1290 ------------------------------------------------------------ 1289

Db       1878 QIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARLRTVMGPKVQGALHLHTLTREA 1937

Qy       1290 ------------------------------------------------------------ 1289

Db       1938 PLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAHGLPALSIDWGMFTEVGMAVA 1997

Qy       1290 -----------------------------------------------------PST------- 1292
                                                                   |:|
Db       1998 QENRGARLISRGMRGITPDEGLSALARLLEGDRVQTGVIPITPRQWVEFYPATAASRRLS 2057

Qy       1293 --------------GDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALF 1338
                            ||||||::|||| |:||| |: :| || ||| || |: :|
Db       2058 RLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQVSHVLRLPEDKIEVDAPL 2117

Qy       1339 TRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL-DALATAL 1389
              : :||||||::||||||||:| :          | | :| : : || |||| |
Db       2118 SSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWLLDDALAVRL 2169
```

RESULT 5

```
Q9KIZ7
ID   Q9KIZ7      PRELIMINARY;      PRT;  7257 AA.
AC   Q9KIZ7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   EpoD.
GN   EPOD.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SMP44;
RX   MEDLINE=20293058; PubMed=10831849;
RA   Julien B., Shah S., Ziermann R., Goldman R., Katz L., Khosla C.;
RT   "Isolation and characterization of the epothilone biosynthetic gene
RT   cluster from Sorangium cellulosum.";
RL   Gene 249:153-160(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SMP44;
RX   MEDLINE=20115953; PubMed=10649995;
RA   Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C., Julien B.;
RT   "Cloning and heterologous expression of the epothilone gene cluster.";
RL   Science 287:640-642(2000).
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
DR   EMBL; AF217189; AAF62883.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR000209; Peptidase_S8.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00107; adh_zinc; 2.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
KW   Oxidoreductase; Phosphopantetheine; Transferase.
SQ   SEQUENCE   7257 AA;  764877 MW;  3A536509FA2FCE99 CRC64;

  Query Match             46.1%;  Score 3322.5;  DB 2;  Length 7257;
  Best Local Similarity   41.2%;  Pred. No. 1.5e-153;
  Matches  786;  Conservative  161;  Mismatches  431;  Indels  531;  Gaps   25;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKT 68
              |:::||||||||:|| |||| ||  :| ||           |||:||||| |:||||: ||:|
Db       3021 ASDEPIAIVGAACRFPGGVEDLESYWQLLAEGVVVSAEVPADRWDAADWYDPDPEIPGRT 3080

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETG 128
               ||: :|| |:  ||:|| ||||||: :|| ||||||| |||||:| |||  |  : ||
Db       3081 YVTKGAFLRDLQRLDATFFRISPREAMSLDPQQRLLLEVSWEALESAGIAPDTLRDSPTG 3140

Qy        129 VFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSS 188
              ||:| ||:||      :            :|| | | || |||:|: ||| || :|:||| |||
Db       3141 VFVGAGPNEYYTQRLRGFTDGAAGLYGGTGNMLSVAAGRLSFFLGLHGPTLAMDTACSSS 3200

Qy        189 LVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR 248
              |||:||||||| |||  || |||:::|:| | ||: ||: |||| ||||:|||: |
Db       3201 LVALHLACQSLRLGECDQALVGGVNVLLAPETFVLLSRMRALSPDGRCKTFSADADGYAR 3260

Qy        249 GEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAG 308
              |||||||||||  |:  || |||:||||||:||||| |||||||||  :|: :|::||: ||
Db       3261 GEGCAVVVLKRLRDAQRAGDSILALIRGSAVNHDGPSSGLTVPNGPAQQALLRQALSQAG 3320

Qy        309 CAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
               :    | :|| ||||| |||||:|||: ||| ||     ||::|:||  |: |  | |||:
Db       3321 VSPVDVDFVECHGTGTALGDPIEVQALSEVYGPGRSEDRPLVLGAVKANVAHLEAASGLA 3380

Qy        369 GLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMS 428
              ||| ||:|:| ||||       ||| :  |  | :  | |    |:|     |||||||:||:|
Db       3381 SLLKAVLALRHEQIPAQPELGELNPHLPWNTLPVAVPRKAVPWGRGARPRRAGVSAFGLS 3440

Qy        429 GTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDV 488
              ||| ||||||||       | || ||  ||:||||::|:|||  | ||  ||   :|    ||||
Db       3441 GTNVHVVLEEAPEVELVPAAPARPVELVVLSAKSAAALDAAAERLSAHLSAHPELSLGDV 3500

Qy        489 AFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQ 548
```

```
Db      3501 AFSLATTRSPMEHRLAIATTSREALRGALDAAAQRQTPQGAVRGKAVSSRGKLAFLFTGQ 3560

Qy       549 GAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQ 608
Db      3561 GAQMPGMGRGLYEAWPAFREAFDRCVALFDRELDQPLREVMWAAPGLAQAARLDQTAYAQ 3620

Qy       609 PALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQAL 668
Db      3621 PALFALEYALAALWRSWGVEPHVLLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMQAL 3680

Qy       669 PAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGAR 728
Db      3681 PAGGAMVAIAASEAEVAASVAPHAATVSIAAVNGPDAVVIAGAEVQVLALGATFAARGIR 3740

Qy       729 TKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRH 788
Db      3741 TKRLAVSHAFHSPLMDPMLEDFQRVAATIAYRAPDRPVVSNVTGHVAGPEIATPEYWVRH 3800

Qy       789 AREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATV 848
Db      3801 VRSAVRFGDGAKALHAAGAATFVEIGPKPVLLGLLPACLGEADAVLVPSLRADRSECEVV 3860

Qy       849 LEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDA------------ 896
Db      3861 LAALGTWYAWGGALDWKGVFPDGARRVALPMYPWQRERHWMDLTPRSAAPAGIAGRWPLA 3920

Qy       897 ------------------------------------------------------------ 896
Db      3921 GVGLCMPGAVLHHVLSIGPRHQPFLGDHLVFGKVVVPGAFHVAVILSIAAERWPERAIEL 3980

Qy       897 -------------------------ARGD---------------RR------------- 902
Db      3981 TGVEFLKAIAMEPDQEVELHAVLTPEAAGDGYLFELATLAAPETERRWTTHARGRVQPTD 4040

Qy       903 -APGA------------------------------------------------------- 906
Db      4041 GAPGALPRLEVLEDRAIQPLDFAGFLDRLSAVRIGWGPLWRWLQDGRVGDEASLATLVPT 4100

Qy       907 ---GHD---------------------EVEEGGA-------------------VR-GGD 921
Db      4101 YPNAHDVAPLHPILLDNGFAVSLLATRSEPEDDGTPPLPFAVERVRWWRAPVGRVRCGGV 4160

Qy       922 RRSA-------------------------------------------RLDHPPP--- 932
Db      4161 PRSQAFGVSSFVLVDETGEVVAEVEGFVCRRAPREVFLRQESGASTAALYRLDWPEAPLP 4220

Qy       933 ----------------------------------------------------------ESGR 936
Db      4221 DAPAERMEESWVVVAAPGSEMAAALATRLNRCVLAEPKGLEAALAGVSPAGVICLWEPGA 4280

Qy       937 REKVEAAGDR-------------------------------------------------- 946
Db      4281 HEEAPAAAQRVATEGLSVVQALRDRAVRLWWVTTGAVAVEAGERVQVATAPVWGLGRTVM 4340

Qy       947 ------------------------------------------------------------ 946
Db      4341 QERPELSCTLVDLEPEVDAARSADVLLRELGRADDETQVVFRSGERRVARLVKATTPEGL 4400

Qy       947 ------PFRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDD 1000
Db      4401 LVPDAESYRLEAGQKGTLDQLRLAPAQRRAPGPGEVEIKVTASGLNFRTVLAVLGM---- 4456

Qy      1001 LPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQA 1060
Db      4457 YPGDAGP---MGGDCAGIVTAVGQGVHHLSVGDAVMTL--GTLHRFVTVDARLVVRQPAG 4511

Qy      1061 LSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATA 1120
Db      4512 LTPAQAATVPVAFLTAWLALHDLGNLRRGERVLIHAAAGGVGMAAVQIARWIGAEVFATA 4571

Qy      1121 GTPEKRAYLESLGV--RYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLR 1178
Db      4572 -SPSKWAAVQAMGVPRTHIASSRTLEFAETFRQVTGGRGVDVVLNALAGEFVDASLSLLT 4630

Qy      1179 SHGRFVELGKRDCYADNQ--LGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIA 1236
Db      4631 TGGRFLEMGKTD-IRDRAAVAAAHPGVRYRVFDILEL------APDRTREILERVVEGFA 4683

Qy      1237 AGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPT-----HAGAGP- 1290
Db      4684 AGHLRALPVHAFAITKAEAAFRFMAQARHQGKVVL-LPAPSAAPLAPTGTVLLTGGLGAL 4742

Qy      1291 ----------------STGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVG 1333
Db      4743 GLHVARWLAQQGAPHMVLTGRRG-LDTPGAAKAVAEIEALGARVTIAASDVADRNALEAV 4801

Qy      1334 AEAL---------------------------FTRLGMDSLMAVELRNRIEASLKLKLST 1365
Db      4802 LQAIPAEWPLQGVIHAAGALDDGVLDEQTTDRFSRVLAPKVTGAWNLHELTAGNDLAFFV 4861

Qy      1366 TFLSTS--------PNIALLAQNLLDALATALSLERVAAENLRAGVQND 1406
Db      4862 LFSSMSGLLGSAGQSNYA-AANTFLDALAAHRRAEGLAAQSLAWGPWSD 4909
```

```
RESULT 6
Q9L8C7
ID   Q9L8C7      PRELIMINARY;      PRT;  7257 AA.
AC   Q9L8C7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Polyketide synthase.
GN   EPOC.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=So ce90;
RX   MEDLINE=20130945; PubMed=10662695;
RA   Molnar I., Schupp T., Ono M., Zirkle R.E., Milnamow M.,
RA   Nowak-Thompson B., Engel N., Toupet C., Stratmann A., Cyr D.D.,
RA   Gorlach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RT   "The biosynthetic gene cluster for the microtubule-stabilizing agents
RT   epothilones A and B from Sorangium cellulosum So ce90.";
RL   Chem. Biol. 7:97-109(2000).
DR   EMBL; AF210843; AAF26921.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR000209; Peptidase_S8.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR002364; QOR_zeta_crystal.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00107; adh_zinc; 2.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   7257 AA;  764272 MW;  621A6820FA57ACE9 CRC64;

  Query Match             45.9%;  Score 3311.5;  DB 2;  Length 7257;
  Best Local Similarity   41.1%;  Pred. No. 5.2e-153;
  Matches  784;  Conservative  162;  Mismatches  432;  Indels  531;  Gaps   25;
```

```
Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKT 68
              |:::|||||||:|| |||| ||  :| ||        |||:||||| |:||||: ||:|
Db       3021 ASDEPIAIVGAACRFPGGVEDLESYWQLLAEGVVVSAEVPADRWDAADWYDPDPEIPGRT 3080

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETG 128
               ||: :|| |:     ||:|| |||||: :||  |||||| |||||:| |||   |  : ||
Db       3081 YVTKGAFLRDLQRLDATFFRISPREAMSLDPQQRLLLEVSWEALESAGIAPDTLRDSPTG 3140

Qy        129 VFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSS 188
              ||:| ||:||        :        :|| |  | ||  |||:|: ||| || :|:||| |||
Db       3141 VFVGAGPNEYYTQRLRGFTDGAAGLYGGTGNMLSVTAGRLSFFLGLHGPTLAMDTACSSS 3200

Qy        189 LVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR 248
              |||:|||||||| |||   || |||:::|:|  |  | ||:  |||:  ||||| |||:|||: |
Db       3201 LVALHLACQSLRLGECDQALVGGVNVLLAPETFVLLSRMRALSPDGRCKTFSADADGYAR 3260

Qy        249 GEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAG 308
              |||||||||||   |:   ||  |||:|||||:||||  ||||||||  :|:  :|::||: ||
Db       3261 GEGCAVVVLKRLRDAQRAGDSILALIRGSAVNHDGPSSGLTVPNGPAQQALLRQALSQAG 3320

Qy        309 CAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
               :      | :|| ||||| |||||||:|||: ||| ||       |||::|:  | |:  |  |  |||:
Db       3321 VSPVDVDFVECHGTGTALGDPIEVQALSEVYGPGRSGDRPLVLGAAKANVAHLEAASGLA 3380

Qy        369 GLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMS 428
               ||| ||:|:| ||||       |||  :  |    |  :  | |    ||        |||||||:||:|
Db       3381 SLLKAVLALRHEQIPAQPELGELNPHLPWNTLPVAVPRKAVPWGRGARPRRAGVSAFGLS 3440

Qy        429 GTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDV 488
              ||| ||||||||          | || || ||:||||::|:||||  ||||  ||   :|      ||||
Db       3441 GTNVHVVLEEAPEVEPAPAAPARPVELVVLSAKSAAALDAAAARLSAHLSAHPELSLGDV 3500

Qy        489 AFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQ 548
              |||||||| ||||||:| ||||||  ||||  ||||||||    |||  |||||   |  ||||||||||||
Db       3501 AFSLATTRSPMEHRLAIATTSREALRGALDAAAQQKTPQGAVRGKAVSSRGKLAFLFTGQ 3560

Qy        549 GAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQ 608
              |||  ||||||:  ||||||| ||  ||::|:|:|||||||  |       || |||||: |
Db       3561 GAQMPGMGRGLYETWPAFREAFDRCVALFDREIDQPLREVMWAAPGLAQAARLDQTAYAQ 3620

Qy        609 PALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQAL 668
              ||||  ||||||||||||||| :: |||||||||||||||||||| ||||||||||||
Db       3621 PALFALEYALAALWRSWGVEPHVLLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMQAL 3680

Qy        669 PAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGAR 728
              |||||||:| | ||:|||:||||||:||||||| || ||||||     | |:  |   ||||   |
Db       3681 PAGGAMVAIAASEAEVAASVAPHAATVSIAAVNGPDAVVIAGAEVQVLALGATFAARGIR 3740

Qy        729 TKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRH 788
              || | |||||||||| |||| | |||  :::|| |       :|||::|       |:::| |||||
Db       3741 TKRLAVSHAFHSPLMDPMLEDFQRVAATIAYRAPDRPVVSNVTGHVAGPEIATPEYWVRH 3800

Qy        789 AREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATV 848
               |  ||| || |||||||| ||||||||  ||||:|||: :|   |: | || | |   |
Db       3801 VRSAVRFGDGAKALHAAGAATFVEVGPKPVLLGLLPACLGEADAVLVPSLRADRSECEVV 3860

Qy        849 LEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDA------------ 896
               | |||  :| || :  | |:|| | ||| || |||||||:|:|     |
Db       3861 LAALGAWYAWGGALDWKGVFPDGARRVALPMYPWQRERHWMDLTPRSAAPAGIAGRWPLA 3920

Qy        897 ------------------------------------------------------------ 896

Db       3921 GVGLCMPGAVLHHVLSIGPRHQPFLGDHLVFGKVVVPGAFHVAVILSIAAERWPERAIEL 3980

Qy        897 -------------------------ARGD---------------RR------------- 902
                                        | ||               ||
Db       3981 TGVEFLKAIAMEPDQEVELHAVLTPEAAGDGYLFELATLAAPETERRWTTHARGRVQPTD 4040

Qy        903 -APGA------------------------------------------------------- 906
               ||||
Db       4041 GAPGALPRLEVLEDRAIQPLDFAGFLDRLSAVRIGWGPLWRWLQDGRVGDEASLATLVPT 4100
```

```
Qy        907 ---GHD---------------------EVEEGGA------------------VR-GGD 921
                 ||                     | |: |                  || ||
Db       4101 YPNAHDVAPLHPILLDNGFAVSLLSTRSEPEDDGTPPLPFAVERVRWWRAPVGRVRCGGV 4160

Qy        922 RRSA-------------------------------------------RLDHPPP--- 932
              ||                                           ||| |
Db       4161 PRSQAFGVSSFVLVDETGEVVAEVEGFVCRRAPREVFLRQESGASTAALYRLDWPEAPLP 4220

Qy        933 ------------------------------------------------------ESGR 936
                                                                    |:|
Db       4221 DAPAERIEESWVVVAAPGSEMAAALATRLNRCVLAEPKGLEAALAGVSPAGVICLWEAGA 4280

Qy        937 REKVEAAGDR-------------------------------------------------- 946
               |: ||  |
Db       4281 HEEAPAAAQRVATEGLSVVQALRDRAVRLWWVTMGAVAVEAGERVQVATAPVWGLGRTVM 4340

Qy        947 ------------------------------------------------------------ 946

Db       4341 QERPELSCTLVDLEPEADAARSADVLLRELGRADDETQVAFRSGKRRVARLVKATTPEGL 4400

Qy        947 ------PFRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDD 1000
                    :|||   :   | || |  |      :|||||  |||||| | |:||:|   |    |||
Db       4401 LVPDAESYRLEAGQKGTLDQLRLAPAQRRAPGPGEVEIKVTASGLNFRTVLAVLGM---- 4456

Qy       1001 LPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQA 1060
               ||   |   :||:|||    |||:||  : ||   |: |  |        ||   | ||: :|
Db       4457 YPGDAGP---MGGDCAGVATAVGQGVRHVAVGDAVMTL--GTLHRFVTVDARLVVRQPAG 4511

Qy       1061 LSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATA 1120
              |:  :|| :|||:||||  ||    :  |:  |||||||| ||||:||||  |: :|||| |||
Db       4512 LTPAQAATVPVAFLTAWLALHDLGNLRRGERVLIHAAAGGVGMAAVQIARWIGAEVFATA 4571

Qy       1121 GTPEKRAYLESLGV--RYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLR 1178
               :| |  |  ::::||    ::: ||:  |     |   ||| |||||||:|:|| :| |  :||
Db       4572 -SPSKWAAVQAMGVPRTHIASSRTLEFAETFRQVTGGRGVDVVLNALAGEFVDASLSLLS 4630

Qy       1179 SHGRFVELGKRDCYADNQ--LGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIA 1236
              : |||:|:|| |          | |   :::|            | |  |  :| ||  ::  |
Db       4631 TGGRFLEMGKTD-IRDRAAVAAAHPGVRYRVFDILEL------APDRTREILERVVEGFA 4683

Qy       1237 AGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPT-----HAGAGP- 1290
               ||      |:     | :    ||| ||||:| ||:||  |   |          ||        ||
Db       4684 AGHLRALPVHAFAITKAEAAFRFMAQARHQGKVVL-LPAPSAAPLAPTGTVLLTGGLGAL 4742

Qy       1291 ----------------STGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVG 1333
                              || |  ||  :|  |   || |  :   ||        ::
Db       4743 GLHVARWLAQQGVPHMVLTGRRG-LDTPGAAKAVAEIEALGARVTIAASDVADRNALEAV 4801

Qy       1334 AEAL---------------------------FTRLGMDSLMAVELRNRIEASLKLKLST 1365
              :|:                            |:|:     :          : :  |   |
Db       4802 LQAIPAEWPLQGVIHAAGALDDGVLDEQTTDRFSRVLAPKVTGAWNLHELTAGNDLAFFV 4861

Qy       1366 TFLSTS--------PNIALLAQNLLDALATALSLERVAAENLRAGVQND 1406
               | | |            | |  |      |||||      | :||::|    |  :|
Db       4862 LFSSMSGLLGSAGQSNYA-AANTFLDALAAHRRAEGLAAQSLAWGPWSD 4909
```

RESULT 7
Q8RJY1
ID   Q8RJY1   PRELIMINARY;   PRT;   2218 AA.
AC   Q8RJY1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   StiF protein.
GN   STIF.
OS   Stigmatella aurantiaca.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Cystobacterineae; Cystobacteraceae; Stigmatella.
OX   NCBI_TaxID=41;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sg a15;

```
RA   Gaitatzis N., Silakowski B., Kunze B., Nordsiek G., Blocker H.,
RA   Hofle G., Muller R.;
RT   "The biosynthesis of the aromatic myxobacterial electron transport
RT   inhibitor stigmatellin is directed by a novel type of modular
RT   polyketide synthase.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ421825; CAD19090.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   2218 AA;  237399 MW;  B85339D5457FA4F5 CRC64;

  Query Match             36.5%;  Score 2634;  DB 2;  Length 2218;
  Best Local Similarity   35.4%;  Pred. No. 1.7e-120;
  Matches  642;  Conservative  182;  Mismatches  447;  Indels  542;  Gaps   25;

Qy          8 RAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGK 67
Db         24 RLAPEPIAIVGMACRYPGGVKDSDSYWDLLKRGGDAVTEIPAARWDINAFYDPNQDSPGK 83

Qy         68 TPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTET 127
Db         84 TTTRWGGFLDDVEGFDPHLFGISPREAASMDPHQRLFLEVAWEALENAGQQTDRLAGSKT 143

Qy        128 GVFIGIGPSEYE-AALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYS 186
Db        144 GVFVGLYTSDYSPVVLSDLNA---INVYTGTGTANNVAAGRLSYLLDLQGPSFVVDTACS 200

Qy        187 SSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGF 246
Db        201 SSLVAVHLACQSLLNRECRMAVVGGVTLLLTPLTLVMASKMQLMAADGRCKTFDSRADGI 260

Qy        247 GRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALAD 306
Db        261 AMGEGCGVVILKRLSDAQADGDRILAVIKASAINQDGRTNGLTAPSGASQQVLLRQALTR 320

Qy        307 AGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASG 366
Db        321 AGISPSQVTYVEAHGTGTPLGDPIEVEALAEVIGQPRPSGPPCVLGSVKANLGHTGAAAG 380

Qy        367 ITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFG 426
Db        381 VAGLIKVVLSMQNEAIPPQLHYRELNSNISLENTSLVIPKELHPWPAGKGPRIAGISAFG 440

Qy        427 MSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQCLG 486
Db        441 WSGTNAHVLVEEAPAAPKARAAVERPVHLLTLSARTEAALREFAERVKTYVSAHADLPIA 500

Qy        487 DVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK-LAFLF 545
Db        501 DLCYSANTGWSKFEHHLAVPAGTTAQLSERLSAFLEGQPQAGTSYGQVTGRRGPGVAFLF 560

Qy        546 TGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTA 605
Db        561 TGQGSQYAGMGRQLYQTQPTFRRALDRCDELLRPLLGRSILSVIY--PEAGQASPIDETA 618

Qy        606 FTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLM 665
Db        619 FTQPALFAFEWALAELWRSWGVVPTVVMGHSVGEYVAACVAGVFTLEEGLALIVERARLM 678

Qy        666 QALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAAR 725
Db        679 QSLPPGGEMAAVFTDGARTAQAIAPYASQVSIAAFNGPSETVISGDGKAVEAILGALSAE 738

Qy        726 GARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYW 785
Db        739 GVKVRRLGVSHAFHSPLMDPMLDAFERAAAKTRFMPPKITLISNLTG-GPSEEFSARAL- 796

Qy        786 VRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEP 845
Db        797 RRHAREPVRFSDGMAALKARGVSVAVEIGPNATLLGIGSRCLPEGSVAWLPSLRKDKDEW 856

Qy        846 ATVLEALGGLWAVGGLVSWAGL---FPSGGRRVPLPTYPWQRERY--------------- 887
Db        857 EVLLATLGALSVRGVPVNWAGFDSDYPR--RRVSLPNYPFQHERFALKQMDGRGVAPSGQ 914

Qy        888 ------------------------------------------------------------ 887
Db        915 VVGHPLLGSRLRVATVDGTLFESKFRADEPSFFSEHRVYGMSTVPATAYLEGALAAAEEI 974

Qy        888 ------------------------------------------------------------ 887
Db        975 FGAGEHALESVDIQEALVLSDAPRTVQYLIGSREEGGTARFQVFSLASSGSGSEGVWTTH 1034

Qy        888 -----------------------------------------WID-------------- 890
Db       1035 SSGQIRIASGAEPLSGTPLTPEQALEAIRTRCPRGLEPRAFYDWLEGQGLEYGPRFRGVQ 1094

Qy        891 ---------------TKADDAARGDRRAPG----------------AGHDEVE------ 912
Db       1095 QVWRGDGEALGLVELPESAVDEAKAFKVHPGLMDACVQLFGVIEYREGAQDDTEVYLPVS 1154

Qy        913 ------------------------------------------------EG---------- 914
Db       1155 IDRYRLKGRLGSKVWSHARIRPSEGTAREТLKADIRLYDEAGRIVAEVEGMCIKRAPREM 1214

Qy        915 ---------------------------GAVRGG------------------DRRSARLD- 928
Db       1215 LARFGQAQFSNWLHEVEWKEEVRSAAAGAARPGSCRWVVFADRSGVAEALIDRLTARGEP 1274

Qy        929 -------------------HP-----------------------------PPESGR 936
Db       1275 VVRVLADDHYSREGDLVRLHPEQPEHVRSLMESAASSGGAVTRYLYLWGLDSSGPVSRGE 1334

Qy        937 REKVE--AAGDRP----------------------------------------------- 947
Db       1335 ASRVDTRAIGSNPLLHLVQELAGQQGSHQLFLVTRGAQATTREHAPIDAFQASLWGFART 1394

Qy        948 ------------------------------------------------------------ 947
Db       1395 IAFEHPELRPVRIDLEPSRKGGSPRETEILLGEVTRSDAASAEDAIAFRGSLRLVSRLMR 1454

Qy        948 ----------------FRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFND 989
Db       1455 QAKRASSNRKLRIPPAPAFRLEISQRGTLEDLTVRPVERRPPGPGEVEIRVRATGLNFRD 1514

Qy        990 VQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIA-LSAGAFATHVT 1048
Db       1515 VLNALGMYPGD-------PGLLGGECAGVISAIGPGVEGFKVGDPVVAVISLGSFATYCT 1567

Qy       1049 TSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQW 1108
Db       1568 VLADFVAHKPAALSFAQAAAIPIAFLTAQYGLQQLGRMAAGSRVLIHAAAGGVGMAAVQL 1627

Qy       1109 AQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGE 1168
Db       1628 AKRAGAEVFATASS-GKWDVLEAMGVEHRMNSRTLDFADEVMARTEGQGIDIVLNSLADD 1686

Qy       1169 LIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALL 1228
Db       1687 FIPKSLSVLRSGGRFLEIGKRGVWTSAQVAER--YPGVSYAIYDLGEVGTKDPSIIRSMF 1744

Qy       1229 EELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGD--PEVQIRIPTHA 1286
Db       1745 RALMAEFEQGTLVPPPLRVFALQDAVEAFRHMALAKHVGKVVITVNDGAPE-SAETTARA 1803

Qy       1287 GAG-PSTGDRDLL 1298
```

```
Db       1804 GAAVPIRGDSTYL 1816

RESULT 8
Q9ADL6
ID   Q9ADL6      PRELIMINARY;      PRT;  6315 AA.
AC   Q9ADL6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Soraphen polyketide synthase A.
GN   SORA.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=So ce26;
RA   Ligon J.M., Hill S., Beck J., Zirkle R., Molnar I., Zawodny J.,
RA   Money S., Schupp T.;
RT   "Characterization of the biosynthetic gene cluster for the antifungal
RT   polyketide soraphen A from Sorangium cellulosum So ce26.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
DR   EMBL; U24241; AAK19883.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00107; adh_zinc; 2.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 4.
KW   Oxidoreductase; Phosphopantetheine; Transferase.
SQ   SEQUENCE   6315 AA;  671301 MW;  7B3FB977006C12FE CRC64;

  Query Match            36.1%;  Score 2606;  DB 2;  Length 6315;
  Best Local Similarity  35.8%;  Pred. No. 1.7e-118;
  Matches  665;  Conservative  193;  Mismatches  489;  Indels  508;  Gaps   32;

Qy         10 AEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKTP 69
Db       2091 SDEPIAIVGMALRLPGGIGDVDALWDFLHQGRDAVEPIPPTRWDAGALYDPDPDAKAKSY 2150

Qy         70 VTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGV 129
Db       2151 VRHAAMLDQVDLFDPGFFGISPREAKHIDPQHRLLLEAAWQALEEAGIVPSTLKDSPTGV 2210

Qy        130 FIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSSL 189
Db       2211 FVGIGASEY---APREPGAEDSEAYIVQGTYASFAAGRLAFTLGLQGPALSVDTACSSSL 2267

Qy        190 VAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGRG 249
Db       2268 VALHLACQALRRDECNLALAAGASVMVSPETFVLLSRLRALAPDGRSKTFSASADGYGRG 2327

Qy        250 EGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAGC 309
Db       2328 EGVIVLALERLRDALAQGRRVLAVVRGTAVNHDGASSGITAPNGTSQKKVLRAALHDARI 2387

Qy        310 AASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITG 369
Db       2388 APADVDVVECHGTGTSLGDPIEVQALAAVYGEGRSAEKPLFLGAVKTNVGHLEAAAGLAG 2447

Qy        370 LLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDW--NTPRRAGVSSFGM 427
Db       2448 VAKIVASLLHNALPPTLHTTPRNPLIAWDALAVAVVDATRPWVRHADGRPRRAGVSAFGL 2507

Qy        428 SGTNAHVVLEEAPAATCTPPAPERPAE-------LLVLSARTASALDAQAARLRDHLETY 480
Db       2508 SGTNAHVILEEAPAIARVEPAASQPASEPLPAAWPVLLSAKSEAAVRAQAKRLRDHLLAK 2567

Qy        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
Db       2568 SELALADVAYSLATTRAHFEQRAALLVKGRDELLSALDSLAQGHSAAVLGRS---GAPGK 2624

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
Db       2625 LAVLFTGQGSQRPTMGRALYDAFPVFRDALDTVAAHLDRDLDRPLRDVLFAPDGSEQAAR 2684

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
Db       2685 LDQTAFTQPALFALEVALFQLLQSFGLKPALLLGHSIGELVAAHVAGVLSLQDACTLVAA 2744

Qy        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
Db       2745 RAKLMQALPQGGAMVTLRASEEEVRDLLQPYDGRASLAALNGPLSTVVAGDEDAVVEIAR 2804

Qy        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTD-EV 779
Db       2805 QAEALGRKTTRLRVSHAFHSPHMDGMLDDFRRVAQSLTYHPARIPIISNVTGARATDHEL 2864

Qy        780 SSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPD----ARPALL 835
Db       2865 ASPDYWVRHVRHTVRFLDGVRALHAEGARVFLELGPHAVLSALAQDALGQDEGTSPCAFL 2924

Qy        836 ASSRAGRDEPATVLEALGGLWAVGGLVSWAGLF-PSGGRRVPLPTYPWQRERYWID-TKA 893
Db       2925 PTLRKGRDDAEAFTAALGALHAAGLTPDWSAFFAPFAPRKVSLPTYAFQRERFWLDASKA 2984

Qy        894 ------------------------------------------------------------ 893
Db       2985 HAADVASAGLTSTDHPLLGAGVPLADRDGFLFTGRLSLSEHPWLADHVVFGTPILPGTAF 3044

Qy        894 ----------------------------------------DDAA-------------- 897
Db       3045 LELALFVAGRVGLDTVEELTLETPLALPSEGALLVQVSVGPLDDAGRRPLSLHSRPQGAP 3104

Qy        898 ------------------------------------------------------------ 897
Db       3105 QDAPWTRHASGSLAPATPSPSFDLHDWPPSGATQVDTQGLYATLESAGLAYGPQFQGLRS 3164

Qy        898 ---RGDR----------------------------------------------------- 901
Db       3165 VWRRGDELFAEAQLPDAAKKDAARFALHPALLDSALHALALDDERAPGVALPFSWGGVSL 3224

Qy        902 ---------------------------------------------------RAPGAGHDE-- 910
Db       3225 RAVGATTLRVRFHRPKGETAGSLVLADAAGGPIASVQALATRITSAEQLRTPGASHHDAL 3284

Qy        911 ---------------------------------------------------VEEGGAVRG--- 919
Db       3285 FRVDWSELPSPTSPSGAPSAVLLGIGGLDLAPEVPLARVADLAALQSALDQGASPPGLVV 3344

Qy        920 ------------------------------------------------------------ 919
Db       3345 VPFMARTADDLIQSAHSITARALALLQAWLADERLASSRLVLLTRRAIAARADEDVKDLA 3404

Qy        920 -----GDRRSARLDHP--------------------PPESGRREKVEAAGDRPF----- 948
Db       3405 HAPLWGLARSAQSEHPELPLFLVDLDLSEASQHTLLAALETGERHS-RLRNGKPFIPRLA 3463
```

```
Qy        949 ----------------RLEIDEPGVLDHLVL-RVTERRAP-GLGEVEIAVDAAGLSFNDV 990
Db       3464 NARSKDELIAPDASNWRLHIPTKGNFDALTLVDAPLARAPLAHGQVRVAVHAAAFNFRDV 3523

Qy        991 QLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTS 1050
Db       3524 LDTLGL----YPGDAGP---LGGEGAGIVTEVGPGVSRYTVGDRVMGIFGAACGPTAIAD 3576

Qy       1051 AALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQ 1110
Db       3577 ARMICPIPHAWSFAQAASVPIIYLTAYYGLVDLGHLKPNQRVLIHAAAGGVGTAAVQLAR 3636

Qy       1111 HVGAEVHATAGTPEKRAYLESLGV--RYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGE 1168
Db       3637 HLGAEVFATA-SAGKWSALRALGFDDAHLASSRDLDFEQHFLRSTHGRGVDVVLDCLARE 3695

Qy       1169 LIDKSFNLLRSHGRFVELGKRDCYADNQLGLR-PFLRNLSFSLVDLRGMMLERPARVRAL 1227
Db       3696 FVDASLRLMPSGGRFVEMGKTDIREPDAVGVAYPGVVYRAFDLIE------AGPDRIEQM 3749

Qy       1228 LEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLG---DPEVQIRIPT 1284
Db       3750 LAELLSLFERGALRPPPITSWDIRHAPQAFRALAQARHVGKFVLTIPRPIDPEGTVLI-- 3807

Qy       1285 HAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALFTRLGMD 1344
Db       3808 -------TGGTGTLGALVARHLVARHGAKHLLLTSR--QGAHAP----GAEA--SRTELE 3852

Qy       1345 SLMA-VELR--NRIEASLKLKLSTTFLSTSPNIALL--AQNLLDALATALSLERV 1394
Db       3853 ALGASVTLRACDAADPRALQALLDSIPSAHPLTAVVHAAGALDDGLLGAMSPERI 3907
```

```
RESULT 9
Q8GBX6
ID   Q8GBX6      PRELIMINARY;      PRT;   5017 AA.
AC   Q8GBX6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Polyketide synthase.
GN   SPIH.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Knauber J., Bloecker H., Mueller R., Nordsiek G., Beyer S.;
RT   "The spirangiene synthase from Sorangium cellulosum So ce90 - Module
RT   duplications as strategy behind the evolution of polyketide
RT   synthases.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ505006; CAD43449.1; -.
SQ   SEQUENCE   5017 AA;  525555 MW;  08BED56182669493 CRC64;

  Query Match             33.5%;  Score 2416.5;  DB 2;  Length 5017;
  Best Local Similarity   39.8%;  Pred. No. 2.4e-109;
  Matches  639;  Conservative  186;  Mismatches  521;  Indels  259;  Gaps   41;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKT 68
Db       1586 ASDEPIAIVGVGLRLPGGVDDVDALFRLLEGERDAVAPIPASRWDVGAVFDPDPEAKGKT 1645

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETG 128
Db       1646 YVRHAAMLDRVDLFDAGFFGISPREARHVDPQHRLLLETAWQALEAATIVPASLRDSTTG 1705

Qy        129 VFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSS 188
Db       1706 VFVGAGASDY--AVLQSSAE-DAEAYAAMGTAASFAAGRLAFTLGLQGPALSIDTACSSS 1762

Qy        189 LVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR 248
Db       1763 LVALHLACQSLRQGECDLALAAGVQVMASPEVFVILSRTRALAPDGRSKTFSANADGFGR 1822

Qy        249 GEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAG 308
Db       1823 GEGVVVLALERLRDARANGHPVLAVVRGTAVNHDGASSGITAPNGSSQQKVLRAALDDAR 1882

Qy        309 CAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
Db       1883 LAPTDVDVVECHGTGTSLGDPIEVQALAAVYGEGRAPERPLRLGAVKTNIGHLEFASGLA 1942

Qy        369 GLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDW--NTPRRAGVSSFG 426
Db       1943 GVVKIVASLRHGTLPATIHTNPRNPHIDWDALPVRVVDAAEPWERSADGSPRRAGVSAFG 2002

Qy        427 MSGTNAHVVLEEAPAATCTPPAPERPAELL--------VLSARTASALDAQAARLRDHLE 478
Db       2003 LSGTNAHVIVEEAP----EPDAPKAARGALPLSGALPFVLSAKSDAALRAQAAALRDQLA 2058

Qy        479 TYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSR 538
Db       2059 RTPDAALVDVAAALATTRSQFDHRAAIAASDHGALVAALEGLAAGVAAPGTF--VAKGAA 2116

Qy        539 GKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDA 598
Db       2117 DKLAFLFTGQGAQRAAMGRGLYDAFPVFRDALDAVASHLDRELDRPLRDVLFAPKGSELA 2176

Qy        599 ALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLV 658
Db       2177 SLLDRTEFTQPALFALEVALFRLVEAWGVTPDVLLGHSVGELAAAHVAGVLSLADACTLV 2236

Qy        659 AARGRLMQALPA-GGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHA 717
Db       2237 AARARLMQALPARDGAMVTVHATEAEVLAALEGRDGRAEIAAINAPSSTVIAGDVDAVLR 2296

Qy        718 IAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTD 777
Db       2297 VAAHFEARGRKATRLRVSHAFHSHHMDPMLDAFRRVAEGLTFHPPRIPIVSNVTGRLAAD 2356

Qy        778 EVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDAR--PALL 835
Db       2357 EIRSPDYWVRHVRSAVRFADGINTLEADGVSSFFELGPHGVLSALGESALANAQREVAFV 2416

Qy        836 ASSRAGRDEPATVLEALGGLWAVGGLVSWAGLF-PSGGRRVPLPTYPWQRERYWIDTKAD 894
Db       2417 PALRDGRADVDALTAALSSLHVRGHRVDWAAFFGPFDPQTVALPTYAFQRDRFWLDAPA- 2475

Qy        895 DAARGDRRAPGAGHD-----EVEEG--GAVRGGDRRSARLDH--------PPPESGRREK 939
Db       2476 -GATRDEAAPAAPEDAAFWRAVDAGDVGAL-GATLNASGEDHLSALATLLPALSAWRRAR 2533

Qy        940 VEAAGDRPFRLEIDEPGVLDHLVLRV-----TERRAPGLGEVEIAVDAAGLS---FNDVQ 991
Db       2534 -------------DERSLVDALRYRVVWKPLTPTAARDVAGTWLLVTLAGAEDALARELE 2580

Qy        992 LALGMVPDDLPGKPNPP-----LLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATH 1046
Db       2581 RALTARGAEVITMPIAPAEADRVRLAVRVQGALADAGE-LRGIV---SFAALDETTLATH 2636

Qy       1047 VTTSA--ALVLPRPQALS--AIEAAAMPV----------------AYLTAW--------- 1077
Db       2637 TALPAGLALTLRLVQALGDMGIEAPLWLVTRGAVSTGRSDRLASAAQSMTWGLGRVVGLE 2696

Qy       1078 -------------------YALDRIARL---QPGE------------RVLIHA------- 1096
Db       2697 HPERWGGLVDVADGDTIDARALDRLVTLLASRDGEDQLALRPTGIFARRLVRAPLGDAPP 2756

Qy       1097 -----------ATGGVG--------LAAVQWAQH---VGAEVHATAGTPEKRAYLESLGV 1134
Db       2757 ARTYTARGTSLVTGGTGALGAHVARFLAGRGAEHLVLVSRRGEDAPGAGALRAELEALGA 2816

Qy       1135 RYV------------------SDSRSDRFVADVRAWTGG------------EGVDVV--- 1161
```

```
Db       2817 RVTLAACDVADRASVAALLARLDAQGDPVRAVVHA--GGVVAQAALASTDLDGVQAVVAA 2874

Qy       1162 -------LNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLR 1214

Db       2875 KVRGALHLHELLGDRELDAFVLFASGAGVWGSGQQGAYAAGNAFLD--------ALAEVR 2926

Qy       1215 -----------------GMM-------LERPARVRALLEELLGLIAAGVF-------TP 1242

Db       2927 RAAGLTATSIAWGAWAGGGMLADHADADAEVELRKRGLVPMAPALAIAALAGALDHGETA 2986

Qy       1243 PPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPT-HAGAGPSTGDRDLLDRL 1301

Db       2987 VTVASVDWARFAPAFASARPR----PLLLDLPEAQRAIEAPSAPAGADGASADRALVEAL 3042

Qy       1302 ASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEAL-----FTRLGMDSLMAVELRNRIE 1356

Db       3043 RPLSEADRARHLVELVVAETAAVLGHAD----ASAIDPHKGFFDLGLDSLVAVELLRKLR 3098

Qy       1357 ASLKLKLSTTFLSTSPNIALLAQNLLDALATALSLERVAAENLRA 1401

Db       3099 ARTGVELPATVTFDHPSPHRLATYVREALPLDAPLAPADATTARS 3143
```

RESULT 10
Q93NX8
ID   Q93NX8   PRELIMINARY;   PRT;   5644 AA.
AC   Q93NX8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   AmphJ.
GN   AMPHJ.
OS   Streptomyces nodosus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF357202; AAK73502.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   5644 AA; 585775 MW; CF6A35E3BA0D2658 CRC64;

Query Match 32.7%; Score 2357; DB 2; Length 5644;
Best Local Similarity 33.0%; Pred. No. 2.3e-106;
Matches 632; Conservative 216; Mismatches 509; Indels 558; Gaps 39;

```
Qy         11 EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGKTP 69

Db       1797 DDPVVIVGMGCRFPGGVGSPEDLWDLLASGSDAITGFPADREWESSRLV--AGEAGGVS- 1853

Qy         70 VTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGV 129

Db       1854 -AQGGFLSDIAGFDADFFGISPREALAMDPQQRILLEVTWEAIERAGVDPTALRGSRTGV 1912

Qy        130 FIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSSL 189

Db       1913 FMGVNGQDYSSLV--MGSRDDVAGHATAGLAVSVVSGRLSYALGLEGPALSVDTACSSSL 1970

Qy        190 VAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGRG 249

Db       1971 VSLHLAAQALRSGECSMALVGGVTVMTTPANFAGFSRMGGLAQDGRCKAFSDSADGTGWS 2030

Qy        250 EGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAGC 309

Db       2031 EGAAVLVVERLSDARRAGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGL 2090

Qy        310 AASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITG 369

Db       2091 RPGDVDAVEAHGTGTPPGDPIEAQALLATYGSDRDPQQPLLLGSVKSNIGHTQAAAGVAG 2150

Qy        370 LLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSG 429

Db       2151 LVKMVMAMRNGVLPRTLHITEPSTHVDWSLGAVQVLTEETAWPETGRVRRAGVSSFGISG 2210

Qy        430 TNAHVVLEEAPAATCTPPAPERPAELLV---LSARTASALDAQAARLRDHLETYPSQCLG 486

Db       2211 TNAHVILEGAPDEPVPAPVADRPVPGAVAWPVSAKSEGALDDQAERLRESADALPAL--- 2267

Qy        487 DVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFT 546

Db       2268 DTAYTLATGRADFEHR------------AVLLAADG-TLTEVARGVAEPHRS--AFLFS 2311

Qy        547 GQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAF 606

Db       2312 GQGAQRLGMGRELHARFPVFAEAFDSVTALLESELDTSVREVMW----GTDEGALNATAF 2367

Qy        607 TQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQ 666

Db       2368 TQPALFAVEVALYRLVESWGVTPDFVAGHSVGEIAAAHVAGVFSLEDACRLVAARAGLMQ 2427

Qy        667 ALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARG 726

Db       2428 ALPSGGAMVAVEATEDEVLPLL---TEGVAVAAVNGPTSVVVSGEEQATLAVAEQLAAQG 2484

Qy        727 ARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEV-SSPGYW 785

Db       2485 RRTSRLRVSHAFHSPLMDPMLEDFRAVAETLSYHEPRIPVVSNLTGEVAAAGVHTHPDYW 2544

Qy        786 VRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEP 845

Db       2545 VRHVREAVRFADGVRGLADRGVTALLEIGPDGVLSALAAASLTDTDTVVVPALRKDRDET 2604

Qy        846 ATVLEALGGLWAVGGLVSW-AGLFPSGGRRVPLPTYPWQRERYW--------------- 888

Db       2605 VSVLSGVARLYVAGVDVDWSAPLSGAGARIADVPTYAFQHERYWPKAAPAALDATGLGLA 2664

Qy        889 ------------------------------------------------------------ 888

Db       2665 SADHPLLGAAMSVAGSDELLLTGSLSAATHPWLADHVVGGMIFFPGTGFLELAVRAADQA 2724

Qy        889 ------------------------------------------------------------ 888

Db       2725 GCDRVEELMIAAPLVLPATGAVQVQISVGAADEEGSRELRFFTRPGEDFDAEWTQHATGR 2784

Qy        889 -------IDTKA---------------------------------------DD------ 895

Db       2785 IGSGEQVIDFDATVWPPRDAEAIDIDGMFERYAADGLEYGPVFRGLRAVWRQDDTVYAEV 2844

Qy        896 ----------------------------AARGDRRA---------------------- 903

Db       2845 ALPESVEDADAFGLHPALFDAALHSTVFLSAEGDTRSLLPFAWEGVSLHADGASTLRVRI 2904

Qy        904 ---------------------------------PG------------------------ 905

Db       2905 ASCGEDTVQIAAVDPGGQPVVSVESLTLRAAGPGDAAEPRRDDSNSLLRVDWTARTLGAP 2964
```

```
Qy        906 ------------------------------------------------------------ 905

Db       2965 AAPATWAVLGEDPFGLTAALAGDSEAVAGVHAPAATIEELAARSGAVPDMVAVTVRGDAD 3024

Qy        906 AGHDEVEE--------------------------GGAVRGGDR--------------RS 924
              || |:  |                          ||   |:|               ||
Db       3025 AGPDDAHELAHEVLALVQGWLAEPAFASSRLVVVTRNAVADGERGAVDLAAAPVWGLVRS 3084

Qy        925 ARLDHP---------------------------PPES---------GRREKVEAA---- 943
              |: ::|                           |::          ||  ::|:
Db       3085 AQSENPGRLLLADVDDTADSLARLPLLAGLFDAEEPQAVVREGTVRVGRLARLESGTSLV 3144

Qy        944 --GDRPFRLEIDEPGVLDHL-VLRVTERRAPGLG-EVEIAVDAAGLSFNDVQLALGMVPD 999
                  |:||    | || | :|   |   |   | || | : ||||:| ||  |||| |
Db       3145 PPAGTPWRLGCRAKGSLDGLALLPYPEAVTPLTGREVRIGIRAAGLNFRDVLNALGMYPG 3204

Qy       1000 DLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQ 1059
              :            | | | || :  ||  |||  |  |: :  | |           |:   |
Db       3205 EAG-------LFGSEGAGVVSEVGPDVTGLAPGDRVMGMVFGGFGPLGVADERLLTRVPD 3257

Qy       1060 ALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHAT 1119
                    |   |:::|: :|||::||  :|  |:|||::||||  ||||:|| | |:|||| ||
Db       3258 DWSWETASSVPLVFLTAYHALKDLAGLRPGEKILIHAGAGGVGMAAIQIAHHLGAEVFAT 3317

Qy       1120 AGTPEKRAYLESLGVR--YVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLL 1177
              | :  |    | ||||     ::: ||: |           | :|:|||||:|:|| :| |  ||
Db       3318 A-SEGKWDVLRSLGVADDHIASSRTLDFETAFTEVAGDKGLDVVLNALAGEFVDASMRLL 3376

Qy       1178 RSHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAA 1237
               : |||:|:|| |     :            :::   || ||:    |   : : :| :|: |
Db       3377 GTGGRFLEMGKTDIRDSDAAS-----DGITYRFFDL-GMV--DPDHIQQMLLDLVDLFER 3428

Qy       1238 GVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLG---DPEVQIRIPTHAGA------ 1288
                |  :| |:     : |   :|||  |: |:|:||:|||:      |||  . : :    |
Db       3429 DVLSPLPVRAWDVRRSREAFRFMSMAKHIGKIVLTMPRAMDPEGTVLVTGGTGGLASALA 3488

Qy       1289 -----------------GP-STGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPE 1329
                               || :  |  ||:  |   || |  :   |  :  :: : :  |
Db       3489 RHLVAEHGIKHLLLTSRRGPDAPGAADLVQALAELGAEARVAACDVADRDALAGLLASVP 3548

Qy       1330 IKVGAEALFTRLGM--DSLMAVELRNRIEASLKLKL--------------------ST 1365
                :      |:     |:   |  ::|      :|::   :: |:               ||
Db       3549 AERPLTAVVHTAGVLDDGILASLTPDRLDTVMRPKVDAAWHLHDLTRDLDLAAFVLYSST 3608

Qy       1366 TFLSTSPNIALLA--QNLLDALATALSLERVAAENL--RAGVQNDFVSSG-ADQD 1415
               : :     ||  |   |      |||||        : | :|   |   |   ::|| :|:|
Db       3609 SGVFGSPGQANYAAGNTFLDALAAHRQSLGLPATSLAWNAWEQGSGMTSGLSDED 3663
```

RESULT 11
Q9L4X2

```
ID   Q9L4X2      PRELIMINARY;      PRT;  5435 AA.
AC   Q9L4X2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   NysJ.
GN   NYSJ.
OS   Streptomyces noursei.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
DR   EMBL; AF263912; AAF71767.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 3.
DR   Pfam; PF00550; pp-binding; 3.
DR   TIGRFAMs; TIGR00128; fabD; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
KW   Oxidoreductase; Phosphopantetheine; Transferase.
SQ   SEQUENCE   5435 AA;  562659 MW;  AA55465DF087A38C CRC64;

  Query Match             32.7%;  Score 2355.5;  DB 2;  Length 5435;
  Best Local Similarity   34.2%;  Pred. No. 2.6e-106;
  Matches  655;  Conservative  183;  Mismatches  513;  Indels  563;  Gaps   47;

Qy          3 DRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPD 61
              :| :   ||::|||:|| :|| || |         | ::      |  :    | :| || |   :
Db         32 NRMLVAAAKEPIAVVGMACRFPGAVDSPEALWEMVATGTDVISGFPDDRGWDLEALRNSG 91

Qy         62 PDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAIAPSA 121
               ||         |  ||   :| ||  ||||||||: |||  ||||     |||:| |  |   :
Db         92 TDARDTDVSQRGGFLDCIADFDPGFFGISPREAVTMDPQQRLLLTTAWEAVERAGIDATT 151

Qy        122 LVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAV 181
               | | || |||    :|    | ::    |   |     | |    | :||:|| ||| || : |
Db        152 LRATRTGAFIGTNGQDYAYLLVRSLDDATGDV--GTGIAASAASGRLSYTLGLEGPALTV 209

Qy        182 DTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSA 241
              ||| ||||||:||| |:|:||:|  |||||::|  :| :|  :|   |:   ||||||||:
Db        210 DTACSSSLVALHLAVQALRNGECGMALAGGVNVMATPGSLVEFSRQGGLARDGRCKAFAD 269

Qy        242 EADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLK 301
               ||| |  ||  |::|:|||  |: :|  :|||:||||:| ||||:  |  || |  ||| |:  |::
Db        270 AADGTGWSEGAGVLLLERLSDAQRNGHPVLAVVRGSAVNQDGASNGFTAPNGPSQQRVIR 329

Qy        302 RALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHP 361
              :|||:|| |·  :  ||||||| ||||||  |:: |  |  |:| | |:|:||:|:|:||
Db        330 QALANAGLATGDIDAVEAHGTGTPLGDPIEAQSILATY--GQDRAHPVLLGSIKSNMGHT 387

Qy        362 EYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISW--GDLRLTVTRARTPWPDWNTPRR 419
              : |||:  |::|:::::||| :| ||     :  :  |   | : :| |  |||:   |||
Db        388 QAASGVAGVIKMIMAMRHGVLPRTLHVDRPSTHVDWTTGSVEL-LTDAH-PWPETGRPRR 445

Qy        420 AGVSSFGMSGTNAHVVLEEAP-----AATCTPPAPERPAELLVLSARTASALDAQAARLR 474
               |:||||:||||||||::|:||      ||  |||    |      ||||| :|| || ||
Db        446 TGISSFGVSGTNAHVIVEQAPDTPAEAADDTPPRTPRTLPWL-LSARTGAALRDQATALL 504

Qy        475 DHLETYPSQCLG----DVAFSLATTRSAMEHRLAVAATSREGL--RAALDAAAQGQTSPG 528
              |||:  |    |     | |||||||||:|:||||||       :|  |  ||  |   |:|
Db        505 DHLDR-PDGDRGPTALDTAFSLATTRAALEHRLAV-VTGTDGTAGRDALTAWLAHGTAPD 562

Qy        529 AVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREV 588
               |    | : | : | ||:|||||| ||||| |:  :  |  |  |  | |  : ||  ||||
Db        563 AHEGHA-AGRTRCAALFSGQGAQRLGMGRELHARFPVFARALDTAVDLLDAELGGTLREV 621

Qy        589 MWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGV 648
              :|      | | |::| |||||||  || |  |  ||||  |: ||||||||: || ||||
Db        622 IW----GTDDAPLNETGFTQPALFAVEVALYRLIESWGVAPDFVAGHSIGEIAAAHVAGV 677

Qy        649 FSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVI 708
              ||||||  |||||  ||||||  |||||::|| | :|:  :        |:|||:| | :|:
```

```
Db        678 FSLEDACTLVAARAGLMQALPRGGAMVAVEATEDEVSPLLTD---GVAIAAINGPTSLVV 734

Qy        709 AGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVS 768
              :|       |:|| :| :| ||  | ||||||||| |||    ||| :||  | | :||
Db        735 SGDETATLAVAARLAEQGRRTTRLRVSHAFHSPLMDPMLAEFRAVAEGLSYGEPQIPVVS 794

Qy        769 NLSGKACTDE-VSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACM 827
              ||:|        : :  ||||| || ||||||::||  || | .|:|:||  ||  |
Db        795 NLTGAVADGTLLGTADYWVRHVREAVRFADGIRALTDAGVGAFLELGPDGTLAALAQQSA 854

Qy        828 PDA--RPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFP-SGGRRVPLPTYPWQR 884
              |||  |  |    |  ||.|    : ||   |  |  |    :  :|    |||| :|
Db        855 PDAVSVPVL----RKDRDEEPAAVAALARLHTAGVPVDWTAFYAGTGAHRTDLPTYAFQY 910

Qy        885 ERYWID------------TKADDAARGD------------------------------ 900
              ||||            | ||    |
Db        911 ERYWPKATYRPADATGLGLTAADHPLLGAAMSVAGSDELLLTGTLSLATHPWLADHVVGG 970

Qy        901 -------------RRAPGAGHDEVEE-------------------------------- 913
                           | |  | | |||
Db        971 MVFFPGTGFLELAVRAADQVGCDRVEELMLAAPLILPATGTVQMQIAVGAADDDGGRDLR 1030

Qy        914 -----------------------------------------------------------G 914
                                                                         |
Db       1031 FFTRPGDDPDAAWAQHATGRITEGERVLALDTTTWPPRDAEPVDIDGLYDRYRANGLDYG 1090

Qy        915 GAVRG-----------------------------------------------GDRRS--- 924
                ||                                               || ||
Db       1091 PVFRGLRAVWRRDTEIYAEVALPEGTADADAFGLHPALFDAVLHSTLFASADGDDRSLLP 1150

Qy        925 -------------------------------------------------------ARLDH 929
                                                                       ||
Db       1151 FAWNGVSLHAAGADALRVRITSCGPDAVEITAVDPQGRPVVSVESLTLRAAGPDAGTADH 1210

Qy        930 ---------------------------------------------PPPESGRREKVEAAG- 944
                                                           |.  :| |: |:| |
Db       1211 RADAGSLFRMDWTPRTVHAPATPATWAVLGTDPIGLTEALTAAGPDTVTGLRDGVDALGE 1270

Qy        945 -----DRPF--------------------------------------------------- 948
                    |||
Db       1271 LTAGDDRPVPDVVAVPLRGATDHGPAGAHDLTRTVLALLQEWLAEERFARSRLLLVTRGA 1330

Qy        949 ---------------------------------------------------RLEIDEP-- 955
                                                                 |: |||
Db       1331 VADGERGPLDLAAAPVWGLVRSAQSENPGRLLLVDLDDTAESAAQLPLLPALLDADEPQA 1390

Qy        956 -------------------------------------GVLDHL-VLRVTERRAPGLG-EVE 977
                                                   | ||| :|    | | |  ||
Db       1391 VVREGTVRVGRLARLDSGRGLVPPPGTPWRLGSRAKGSLDGLALLPHPEARRPLTGHEVR 1450

Qy        978 IAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIA 1037
              : :  ||||:| ||  |||| | |        | | | || :| ||  | ||  |  .|:
Db       1451 VGIRAAGLNFRDVLNALGMYPGDAG-------LFGSEAAGVVVEVGPEVTGLAPGDRVMG 1503

Qy       1038 LSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAA 1097
              :  ||           || |: | |    |     |::|: :|||:|||  :  |: ||:||:||
Db       1504 MLFGGFGPLGIADARLLTPVPADWSWETGASVPLVFLTAYYALKELGGLRAGEKVLVHAG 1563

Qy       1098 TGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVR--YVSDSRSDRFVADVRAWTGG 1155
               ||||:||:| |:||||| |||  :  |  |  ||||   ::: ||:  |       |
Db       1564 AGGVGMAAIQIARHVGAEVFATA-SEGKWDVLRSLGVADDHIASSRTLDFEAAFAEVAGD 1622

Qy       1156 EGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRG 1215
              |:|||||:||||  :| |  ||   |||:|:|| | |  |  :   ||:  |  |||
Db       1623 RGLDVVLNALSGEFVDASMRLLGDGGRFLEMGKTDIRAADSVP-----DGLSYHSFDL-G 1676

Qy       1216 MMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLG- 1274
              |:   :: :| :|:|: |  |       |:  :  :  :  : :|||  |: |||:||:|||:
Db       1677 MV--DPEHIQRMLLDLVELFDRGALAALPVRSWDVRRAGEAFRFMSLAQHIGKIVLTVPQ 1734

Qy       1275 --DPEVQIRIPTHAG--AG-------PSTGDRDLL--DRLASAAPAARA----------- 1310
                ||:  :  :   |  ||      | |  ||  |   ||| | |
Db       1735 PLDPDGTVLLTGGTGGLAGLLARHLVTEHGARHLLLAGRRGPDAPGAAALHAELTALGAE 1794

Qy       1311 ---AALEAFLRTQVSQVLRTPEIK-----------VGAEALFTRLGMDSLMAVELRNRIE 1356
                 || :   || :: :| |   :           | :    | |  |  |  |  | || :::
Db       1795 VTVAACDVADRTALAALLATVPAEHPLTAVVHTAGVLDDGTLTALNPDRLATV-LRPKVD 1853

Qy       1357 AS---------LKLKLSTTFLSTS-----PNIALLA--QNLLDALAT---ALSL 1391
              |:         | |     : ||:      |   |   |    |||||     || |
Db       1854 AAWHLHDLTRHLDLAAFVLYSSTAGVMGGPGQANYAAGNTFLDALAAHRHALGL 1907
```

RESULT 12
Q9L4W3

```
ID   Q9L4W3      PRELIMINARY;      PRT; 11096 AA.
AC   Q9L4W3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   NysC.
GN   NYSC.
OS   Streptomyces noursei.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71776.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   11096 AA;  1150415 MW;  776CAEAFCAE551DD CRC64;

  Query Match             32.3%;  Score 2331.5;  DB 2;  Length 11096;
  Best Local Similarity   32.9%;  Pred. No. 1e-104;
  Matches  629;  Conservative  182;  Mismatches  519;  Indels  583;  Gaps   34;

Qy         10 AEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGKT 68
              |:||| :|| :|| ||||         | |::   | :    |  | ||     :||||:  | :
Db       3499 ADDPIVVVGMACRFPGGVTSPEDLWRLVDDGTDAITTFPTNRGWDLDNLYDPDPEHFGTS 3558

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETG 128
                        || :   || :|||:||||||   |||||  |||:|    |  |  |  | |  |: ||
Db       3559 YTRSGGFLHEAGEFDPAFFGMSPREALATDSQQRLLLESSWEAIERAGIDPLTLRGSATG 3618

Qy        129 VFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSS 188
              || |:   |:|: |          | :    | |:   || :|:|:||||  || |  ||||  |||
Db       3619 VFAGVMYSDYGSIL----GGKEFEGFQGQGSAGSVASGRVSYALGFEGPAVTVDTACSSS 3674

Qy        189 LVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR 248
              |||:| |  |:||:||||  ||||||::| :|||  |   |: | || ||| |||:  |||
Db       3675 LVALHWAAQALRAGECSLALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKAFAEAADGVGW 3734
```

```
Qy        249 GEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAG 308
Db       3735 SEGVGILVLERQSDAVRNGHEILAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGG 3794

Qy        309 CAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
Db       3795 LSTADVDAVEAHGTGTTLGDPIEAQALLATYGRDRDPENPLWLGSLKSNIGHTQAAAGVA 3854

Qy        369 GLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMS 428
Db       3855 GVIKMVMAMRHGVLPQTLHVDAPSSHVDWSVGAVELLTEQTAWPETGRVRRAGVSSFGIS 3914

Qy        429 GTNAHVVLE-----EAPAATCTPPAPERPAEL-LVLSARTASALDAQAARLRDHLETYPS 482
Db       3915 GTNAHVIVEQPALVESPAA---EPSGREPGVVPLPLSGKSPEALRDQAARLLAGLAERPA 3971

Qy        483 QCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLA 542
Db       3972 LRPLDLGYSLATTRSAFDHRAVVLATDRADAVRALTALAAADADLSAV--VGDTRTGRHA 4029

Qy        543 FLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQEL--DRPLREVMWAEPASVDAAL 600
Db       4030 VLFSGQGSQRLGMGRELYERFPVFAEALDVAIDHLDAALPAQASLREVMWGD----DVEL 4085

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
Db       4086 LDETGWTQPALFAVEVALFRLVESWGVRPDFVAGHSIGEIAAAHVVGVFSLEDACRLVAA 4145

Qy        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
Db       4146 RATLMQALPTGGAMIAIQAAEDEVTQHLTD---DVSIAAVNGPTSVVVSGAESAARTVAD 4202

Qy        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACT-DEV 779
Db       4203 RLAENGRKTTRLRVSHAFHSPLMDPMLAEFRAVAEGLSYATPTLPVVSNLTGRLATADDL 4262

Qy        780 SSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMP-DARPALLASS 838
Db       4263 CSAEYWARHVREAVRFADGVSTLENEGVTTFLELGPDGVLSAMAQQSLTGDA--ATVPAL 4320

Qy        839 RAGRDEPATVLEALGGLWAVGGLVSWAGLFP-SGGRRVPLPTYPWQRERYW----IDTK- 892
Db       4321 RKDRDEETSALTALAHLHTAGLRVDWAAFFAGSGATRVDLPTYAFQHATYWPTGTLPTAH 4380

Qy        893 ---------------------------------------ADDAARGDRRAPG------- 905
Db       4381 AAAVGLTAAEHPLLNGSVELAEGEGVLFTGRLSLQSHPWLADHAVMGQVLLPGTALLELA 4440

Qy        906 ------AGHDEVE-----------EGGAV------------------------------ 917
Db       4441 FRAGDEAGCDRVEELTLAAPLVLPERGAVQTQVRVGVADDTGRRTVTVHSRPEHATDVSW 4500

Qy        918 -----------------------------------------------------------R 918
Db       4501 TQHATGTLTMGSAPADTGFDATAWPPADAEPLATDDCYARFTTLGFAYGPVFQGLRAAWR 4560

Qy        919 GGD--------------------------------------------------------- 921
Db       4561 AGDVLYAEVALAESTGDEATAFGLHPALLDAALHASLVAHEGEESNGGLPFSWEGATLYA 4620

Qy        922 ------------------------------------------------------------ 921
Db       4621 TGATALRVRLTPTGTDGRSVAIAVADTAGRPVAAIDNLVSRRVSGDQLTGAAGLARDALF 4680

Qy        922 ------------------------------------------------------------ 921
Db       4681 TLDWNPVPENLVPENPVPENTGGGHAQDQDGRPAAATVALVGADGTAIAADLTAAGIHTT 4740

Qy        922 ------------------------------------------------------------ 921
Db       4741 LHPDLTTLATTDADVPKTVLIPLTGTGTGTGTGTESTDGIGTGAAESDASAPSPAEVAHT 4800

Qy        922 -------------------------------------------------RRSAR------ 926
Db       4801 LSTAALALVQEWTAQERFAGSRLAFVTTGATAAGGTDVMDVAAAAVWGLVRSAQSEAPDT 4860

Qy        927 ---LDHPPPESGRREKVEAA---------------------------------------- 943
Db       4861 FVLIDRDPGPAGTHDRTAAAERGQLLLRALHTDEPQLALRDGGVLAARLARFDTAAALTP 4920

Qy        944 -GDRPFRLEIDEPGVLDHLVLR-VTERRAPGLG-EVEIAVDAAGLSFNDVQLALGMVPDD 1000
Db       4921 PADRAWRLDSTAKGSLNGLALTPYPAALAPLTGHEVRVEVRAAGLNFRDVLNALGMYPGD 4980

Qy       1001 LPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQA 1060
Db       4981 DVGS------FGSEAAGVVVEVGPEVTGLAPGDQVMGMITGSFGSLAVDDARRLARLPED 5034

Qy       1061 LSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATA 1120
Db       5035 WSWETGASVPLVFLTAYYALKELGGLRAGEKVLVHAGAGGVGMAAIQIARHVGAEVFATA 5094

Qy       1121 GTPEKRAYLESLGVR--YVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLR 1178
Db       5095 -SEGKWDVLRSLGVADDHIASSRTLDFEAAFAEVAGDRGLDVVLNSLAGDFVDASMRLLG 5153

Qy       1179 SHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAG 1238
Db       5154 DGGRFLEMGKTDIRAADSVP-----DGLSYQSFDLAWVV---PETIGTMLAELMDLFRTG 5205

Qy       1239 VFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTL---GDPEVQIRIPTHAGAG------ 1289
Db       5206 ALRPLPVRTWDVRHAKDAFRFMSMAKHIGKIVLTLPRSWKPEGTVLVT--GGTGGLGGLV 5263

Qy       1290 -----PSTGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALFTRLGM- 1343
Db       5264 ARHLVRSCGVRHLLLTSRSGVGAAGAAGLVAELES-----------LGARVVVAACDVG 5311

Qy       1344 DSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATALSLERVAA 1396
Db       5312 DGSAVAELVAGVSESYPLSA----------VVHAAGVLDDGVVGSLTPERLAA 5354
```

```
RESULT 13
Q9EWA1
ID   Q9EWA1      PRELIMINARY;      PRT;  9507 AA.
AC   Q9EWA1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   PimS2 protein.
GN   PIMS2.
OS   Streptomyces natalensis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=68242;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20547809; PubMed=11094342;
RA   Aparicio J.F., Fouces R., Mendes M.V., Olivera N., Martin J.F.;
RT   "A complex multienzyme system encoded by five polyketide synthase
RT   genes is involved in the biosynthesis of the 26-membered polyene
RT   macrolide pimaricin in Streptomyces natalensis.";
RL   Chem. Biol. 7:895-905(2000).
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
DR   EMBL; AJ278573; CAC20921.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 6.
```

```
DR   Pfam; PF00106; adh_short; 3.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00037; MYB_1; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
KW   Oxidoreductase; Phosphopantetheine; Transferase.
SQ   SEQUENCE   9507 AA;  994228 MW;  57BD80C8AD37E89F CRC64;

  Query Match             32.3%;  Score 2330;  DB 2;  Length 9507;
  Best Local Similarity   37.6%;  Pred. No. 9.9e-105;
  Matches  624;  Conservative  179;  Mismatches  544;  Indels  312;  Gaps   40;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGK 67
Db       4856 AADEPIAIIGMSCRYPGGVGSPEDLWRLVTEGGDATGEFPADRGWDAEGLYDPDPDRAGH 4915

Qy         68 TPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTET 127
Db       4916 TYSTRGGFLHDAADFDASFFGISPREALAMDPQQRLLLETSWEAMERAGIDPATLRGSST 4975

Qy        128 GVFIGIGPSEYEA--ALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAY 185
Db       4976 GTFIGASYQDYRAFGAGPDGA-----EGHLITGTISSVLSGRLSYTYGFEGPAVSLDTAC 5030

Qy        186 SSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADG 245
Db       5031 SSSLVALHLACQSLRNGESSLALAGGVSIMSTPGAFVGFSRQRAMAADGRCKAYSDQADG 5090

Qy        246 FGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALA 305
Db       5091 MSLAEGVGLVLVERLSDARRNGHQVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALA 5150

Qy        306 DAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYAS 365
Db       5151 NAEVDPDGIDVIDGHGTGTALGDPIEAQALLATYGQGRDPEHPLLLGSVKSNIGHTQMAS 5210

Qy        366 GITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSF 425
Db       5211 GVASVIKMVMAMRHAAVPRSLHIDRPSSHVDWSSGAIQLLTEPLPWPATGHPRRAGVSSF 5270

Qy        426 GMSGTNAHVVLEEAP---AATCTPPAPERPAE----LLVLSARTASALDAQAARLRDHLE 478
Db       5271 GLSGTNVHTILEQAPEDRTEEDTDTAPARPADPTPVPVTLSGRTEAGLRAQAGRLLTHLA 5330

Qy        479 TYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSR 538
Db       5331 DHPELPLTDLAFSLATSRAALEHRAAVVTDDPDTLTRALTALRDATPDGGLLTGRPD--R 5388

Qy        539 GKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDA 598
Db       5389 GRLAFLFTGQGSQRPGMGRELYDRHPVYAEALDAVLARFDLELDRPLREILFADPGTPEA 5448

Qy        599 ALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLV 658
Db       5449 ELLDDTGYTQPALFALEVALFRLAESWGLRPDYVAGHSIGELAAAHAAGVLSLEDACTLV 5508

Qy        659 AARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAI 718
Db       5509 AARGRLMAALPSGGAMASVEATEDEAAAVLAPYEGRAAVAAVNTPTSLVVSGDEDAVHAV 5568

Qy        719 AAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGK-ACTD 777
Db       5569 AAHFHELGRRTKQLRVSHAFHSPHMDAMLDDFTRAAEGITFHAPELPLVSTLTGELILAD 5628

Qy        778 EVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPA-LLA 836
Db       5629 DLCSAPYWTRQVRGTVRFADAVRHLTEHGVTTFFELGPDAVLSGAVRESTGEQTPATAVP 5688

Qy        837 SSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGG-RRVPLPTYPWQRERYWIDTKADD 895
Db       5689 ALRRDRPEAPALTTALARLHLHGVRVDWDAVFAGGGARRIDLPTYPFQRERYWPEAAAGA 5748

Qy        896 AA-----------------RGDRRAPGAGHD-----------------EVEEGGAVRGG 920
Db       5749 AAPAHPADAADAEFWSAVERADLPSLGSSLDLDDDTLTAVVPALSSWRRKSRERSTVDGW 5808

Qy        921 DRR--------SARLDHP-------PPESGRREKVEAA----GDRPFRLEI--------- 952
Db       5809 RYRTTWKPLTGTATAGHPAGTWLVLAPADGDRAWTDAVAGALGADAVRVEVSTAQRQELA 5868

Qy        953 -----------DEPGVL-----------------DHLVLRVTERRAPGLGEVE------ 977
Db       5869 ERLSELAAEHGEFAGVLSLLATAGEGAEDADDATTEGLLLTATALQALGDAGIDAPLWCV 5928

Qy        978 ----IAVDAAGLSFNDVQLA---LGMV-----PDDLPGKPNPPLLGGECAGRIVAVGEG 1025
Db       5929 TRTAVAVDRAEHPARPAQAAVWGLGRVAALEHPQRWGGLVDLPDELDGATLRRLAA---- 5984

Qy       1026 VNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIAR 1085
Db       5985 ---------VLADSGDEDQLAVRTTATFV----RRLAHHRAAPAPEA----------AT 6020

Qy       1086 LQPGERVLIHAATGGVGLAAVQWAQHVGAEVHA---------TAGTPEKRAYLESLGVRY 1136
Db       6021 FRPTGTVLVTGGTGALGGHVARWLAEAGAE-HLLLVSRRGTDAPGAGELAAEIAELGARV 6079

Qy       1137 ---VSDSRSDRFVADVRA------------WTGGEGVDVVLNSLSGE------------ 1168
Db       6080 TVAACDTADRDALAAVLAAIPDEHPLTAVFHTAGTVDDGTLDTLTPEQFASVLRTKVTAT 6139

Qy       1169 --------------------------------------LIDKSFNLLRSH---------G 1181
Db       6140 RNLHETTREQDLSAFVLFSSVAGTLGAPGQGNYAAGNAFLDAFAAHRRAHGLPATSLAWG 6199

Qy       1182 RFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAG--- 1238
Db       6200 PWAETG----MAADGTGIQDRVRRGGFTPMPPR---LALTALRRAIEHDATALTLADIDW 6252

Qy       1239 -----VFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTG 1293
Db       6253 QRYAEVFT----ATRPSLFVAD-------LPELQQVTPPAGTPEAALREPA--------- 6292

Qy       1294 DRDLLDRLASAAPAARAAALEAFLRTQVSQVL-RTPEIKVGAEALFTRLGMDSLMAVELR 1352
Db       6293 ---LRQRLAGLSPAARPRFVLDLVRTQVAAVLGHSGTSAIGAERAFSDLGFDSLTTVELR 6349

Qy       1353 NRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATALS-------LERVAAE-------- 1397
Db       6350 NTLTATTGLKLPATLVYDYPTPTALADFLLAELLGALPESGAPAFVGRAADDDPIVIVGM 6409

Qy       1398 NLR--AGVQN-----DFVSSG--------ADQDWEIIAL 1421
Db       6410 NCRFPGGIRSPEDLWQLLSSGEDAISGFPADRGWDLDAL 6448

RESULT 14
Q93NW6
ID   Q93NW6     PRELIMINARY;      PRT; 10917 AA.
AC   Q93NW6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   AmphC.
GN   AMPHC.
OS   Streptomyces nodosus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=40318;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., Oliynyk M.;
RT   "The amphotericin biosynthetic gene cluster from Streptomyces
```

```
RT   nodosus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- COFACTOR: ZINC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AF357202; AAK73514.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   Pfam; PF00550; pp-binding; 6.
DR   TIGRFAMs; TIGR00128; fabD; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00059; ADH_ZINC; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Metal-binding; Oxidoreductase; Phosphopantetheine; Transferase; Zinc.
SQ   SEQUENCE   10917 AA;  1132905 MW;  15AC5956BB5810A1 CRC64;

  Query Match             32.3%;  Score 2328.5;  DB 2;  Length 10917;
  Best Local Similarity   33.3%;  Pred. No. 1.4e-104;
  Matches  620;  Conservative  192;  Mismatches  518;  Indels  533;  Gaps   32;

Qy         11 EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGKTP 69
Db       3500 DDPIVIVGMACRFPGDVTSPEDLWRLVSEGTDAISDFPANRGWDVDSLYHPDPEHSGTSY 3559

Qy         70 VTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGV 129
Db       3560 TRSGGFLHDAGDFDPDFFGMSPREALATDSQQRLLLETSWEAIERAGIDPLSLRDSRTGV 3619

Qy        130 FIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSSL 189
Db       3620 FAGVMYHDYGSLL----GGKEFEGFQGQGSAGSVASGRVSYTFGFEGPAVTVDTACSSSL 3675

Qy        190 VAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGRG 249
Db       3676 VALHWAAQALRSGECSLALAGGVTVMSTPSTFIEFSRQQGLSADGRSKAFAESADGVGWS 3735

Qy        250 EGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAGC 309
Db       3736 EGVGMLLLERQSDAIRNGHEILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGL 3795

Qy        310 AASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITG 369
Db       3796 STGDVDAVEAHGTGTTLGDPIEAQALIATYGQNRTDEQPLWLGSLKSNIGHTQAAAGVAG 3855

Qy        370 LLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSG 429
Db       3856 VIKMVMSMRHGVLPCTLHVDAPSSHVDWTEGAVELLTEQTEWPETDRVRRAAVSSFGISG 3915

Qy        430 TNAHVVLEE---APA-ATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQCL 485
Db       3916 TNAHVILEQPVVAPAPATADVDGAEPAVVPLILSGKSPEALRDQAARLLDTVREHTALRP 3975

Qy        486 GDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLF 545
Db       3976 LDLGHSLATSRSAFDHRAVVLATGREDALRALTALADDEANSAAVTGRTRS--GRRAALF 4033

Qy        546 TGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTA 605
Db       4034 SGQGSQRLGMGRELYGRFPVFAEALDAVLAVLDGELEGSLREVMWGE----DAGLLNETG 4089

Qy        606 FTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLM 665
Db       4090 WTQPALFAVEVALYRLVESWGVRPDFVAGHSIGEIAAAHIAGVFSLEDAARLVAARGRLM 4149

Qy        666 QALPAGGAMVSIEAPEADVAAAVAPHAA-SVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724
Db       4150 QALPAGGAMVAVQATEDE----VIPYLSDEVSIAALNGPASVVVSGAEGAVLEVAARFEA 4205

Qy        725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACT-DEVSSPG 783
Db       4206 EGRKATMLRVSHAFHSPLMDLMLDEFRAVAETLSFAAPVIPVVSNLTGSLATAEELCSPE 4265

Qy        784 YWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRD 843
Db       4266 YWVRHVREAVRFADGVSALEAQGVTTFLELGPDGVLSAMAQESLTGDEAVTVALLRKDRD 4325

Qy        844 EPATVLEALGGLWAVGGLVSWAGLFP-SGGRRVPLPTYPWQRERYW-------IDTK--- 892
Db       4326 EETAALTALAALHTAGVRLDWNAFFAGTGATRVDLPTYAFQHTTYWPSGTHSAADAAAVG 4385

Qy        893 ----------------------------------ADDAARGDRRAPG----------- 905
Db       4386 LTAAGHPLLNGSVQLAEGEGVLFTGRLSLQSHPWLADHAVHGTVLLPGTALLDLAFRAGD 4445

Qy        906 -AGHDEVE-----------EGGAVR---------------------------------- 918
Db       4446 EVGCDRVEELTLAAPLVLPERGAVQTQVRVGVADDTGRRTVTVHSRPEDATDVSWTQHAT 4505

Qy        919 -----GGDRRSARLD--------------------------------------------- 928
Db       4506 GTLMPGGTARDTGFDATVWPPTGAEALDTEGCYERFTALGFGYGPVFQGLKAAWRSGDVL 4565

Qy        929 -------------------HP--------------------------------------- 930
Db       4566 YAEVVLQEAGAADDEAGFGLHPALLDAALHASLVAHEGEEGNGGLPFAWEGAVLHATGAT 4625

Qy        931 -------PPESGR----------------------------------------------- 936
Db       4626 ALRVRLTPTDEARQSVAIAVCDTAGQPVASIDRLLVRAVTGDQLNEAAGIARDALFALGW 4685

Qy        937 ------------------------------------------------------------ 936
Db       4686 TPLPQDGDIPAAVALVGPDTTGLAADLTAAGAAVTAHQDLAALAADGTDIPGTVVVTVDG 4745

Qy        937 ------------------------------------------------------------ 936
Db       4746 TAAPVAEAAHTVAAEALALAQRWTAEDRFAGARLVFATCGAVEAGGSQVTDVAAAAVWGL 4805

Qy        937 ----------------REKVEAAGD----------------------------------- 945
Db       4806 VRSAQSESPDTFVLVDREAADAAGDAAERGSLFLAALGSGEPQLALRDDVVLAGRLARFD 4865

Qy        946 ----------RPFRLEIDEPGVLDHLVLRV-TERRAPGLG-EVEIAVDAAGLSFNDVQLA 993
Db       4866 AAGAMTLPAERAWHLDSTTPGSINGLALTPHPEALEPLTGHQARIEVRAAGLNFRDVLKA 4925

Qy        994 LGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAAL 1053
Db       4926 LGMYPGD-AGR------YGHEAAGVVVEVGPDVTHIAPGDRVMGMVSGSFASFAVTDARR 4978

Qy       1054 VLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVG 1113
Db       4979 LTHLPEQCTWEIGASLPLVFLTAYHALKELGGLTAGEKVLIHAGAGGVGMAAIQIARHFG 5038

Qy       1114 AEVHATAGTPEKRAYLESLGVR--YVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELID 1171
Db       5039 AEVFATA-SEAKMDVLRSLGVADDHIASSRTLDFEAAFAAVAGENGLDVVLNSLAGEFVD 5097

Qy       1172 KSFNLLRSHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEEL 1231
Db       5098 ASMRLLGTGGRFLEMGKTDIREDDSVPDGITYQSFDLAFVD--------PEVIGAMGREL 5149

Qy       1232 LGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHAGAGPS 1291
Db       5150 TELFAAGDLHPLPVRAWDIRHATDAFRHMSMARHIGKIVLTVP------RAWSSEGTVLV 5203
```

```
Qy       1292 TGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALFTRLGMDSLMAVEL 1351
              ||    |  |      ||   :|  :|   :                 | :| |:|:  | ||
Db       5204 TGGTGGLGGL-----VARHLVMERGVRRLL--------------LTSRSGLDAAGAREL 5243

Qy       1352 RNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATALSLERVAAEN-LR-----AGVQN 1405
               :|              :       ::: | :: |  |     :  : ||: ||    ||| :
Db       5244 VAELE-----------NLGAEVSVAACDVADRDAVDTLIAGIPAEHPLRAVVHTAGVLD 5291

Qy       1406 DFV 1408
              | |
Db       5292 DGV 5294


RESULT 15
O53490
ID   O53490      PRELIMINARY;      PRT;  4151 AA.
AC   O53490;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Polyketide synthase.
GN   PKS12 OR RV2048C OR MT2108 OR MTV018.35C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL021899; CAA17262.1; -.
DR   EMBL; AE007061; AAK46387.1; -.
DR   HSSP; P25715; 1MLA.
DR   TIGR; MT2108; -.
DR   TubercuList; Rv2048c; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR006163; Pp_bind.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; adh_zinc; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW   Phosphopantetheine; Transferase; Complete proteome.
FT   CONFLICT   2147   2147       H -> Q (IN REF. 2).
FT   CONFLICT   2260   2261       DI -> TV (IN REF. 2).
FT   CONFLICT   2268   2268       W -> H (IN REF. 2).
FT   CONFLICT   2272   2272       S -> A (IN REF. 2).
FT   CONFLICT   2279   2280       AF -> PY (IN REF. 2).
FT   CONFLICT   2282   2283       AA -> GR (IN REF. 2).
FT   CONFLICT   2287   2287       T -> V (IN REF. 2).
FT   CONFLICT   2289   2289       F -> W (IN REF. 2).
FT   CONFLICT   3004   3004       S -> L (IN REF. 2).
FT   CONFLICT   3649   3649       P -> A (IN REF. 2).
SQ   SEQUENCE   4151 AA;  431577 MW;  536F644F111D7A5D CRC64;

  Query Match             32.2%;  Score 2321;  DB 16;  Length 4151;
  Best Local Similarity   33.2%;  Pred. No. 8.7e-105;
  Matches  613;  Conservative  203;  Mismatches  535;  Indels  498;  Gaps   31;

Qy          6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDA 64
              :   :|||||||| :|| ||||           | :|       ||  :     ||:| || |  ::||||
Db       2051 VRTTSEDPIAIVGMACRYPGGVNSPDDMWDMLIQGRDVLSEFPADRGWDLAGLYNPDPDA 2110

Qy         65 PGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124
              |           |:  |   || :||: | ||| ||| ||:|||: |||| | | |:  | |
Db       2111 AGACYTRTGGFVDGVGDFDPAFFGVGPSEALAMDPQHRMLLELSWEALERAGIDPTGLRG 2170

Qy        125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184
              : ||||  |:        |           |:  ::          | : || :||::| ||| || |:||||
Db       2171 SATGVFAGVMTQGY-----GMFAAEPVEGFRLTGQLSSVASGRVAYVLGLEGPAVSVDTA 2225

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244
              ||||||:|:|  ||||||  ||||||::  :|      |         |:  | : ||||||:| ||
Db       2226 CSSSLVALHMAVGSLRSGECDLALAGGVTVNATPDIFVEFSRWRGLSPDGRCKAFAAAAD 2285

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304
              |   | ||   ::||:|||  ||   :|||:  |||:  ||||:|||  ||| ||  |: :  ||
Db       2286 GTGFSEGGGMLVLQRLSDARRLGHPVLAVVVGSAVNQDGASNGLTAPNGPSQQRVVRAAL 2345

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGR-DVATPLLIGSVKTNLGHPEY 363
              |:|| :|:  |   ||  |||||||||||  |||  |||   |    :      || :||||:|:||  :
Db       2346 ANAGLSAAEVDVVEGHGTGTTLGDPIEAQALLATYGQDRGEPGEPLWLGSVKSNMGHTQA 2405

Qy        364 ASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVS 423
              |:|: |::|:||:::|  :|| ||       :| :  |        : :   |      ||        |||||
Db       2406 AAGVAGVIKMVLAMRHELLPATLHVDVPSPHVDWSAGAVELLTAPRVWPAGARTRRAGVS 2465

Qy        424 SFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQ 483
              |||:||||||||::|   |                  |   |:||:: |||  |||||    ::
Db       2466 SFGISGTNAHVIIEAVPVPRREAGWAGPVVPWVVSAKSESALRGQAARLAAYVRGDDGL 2525

Qy        484 CLGDVAFSFLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAF 543
               :  | :|||   ||  |||    |     |  :|  ||  |           ||  | ::|  ||   |
Db       2526 DVADVGWSLA-GRSVFEHRAVVVGGDRDRLLAGLDELAGDQLGGSVVRGTA-TAAGKTVF 2583

Qy        544 LFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQ 603
              :| |||:| ||||  ||:   :   | |||:   |     :: | ||||||||       |   ||:
Db       2584 VFPGQGSQWLGMGMGLHAGYPVFAEAFNTVVGELDRHLLRPLREVMWGH----DENLLNS 2639

Qy        604 TAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGR 663
               |  | ||||  |  ||   |   ||||  |:  | ||||||| || ||||  |||:| |  |||||||
Db       2640 TEFAQPALFAVEVALFRLLGSWGVRPDFVMGHSIGELSAAHVAGVLSLENAAVLVAARGR 2699

Qy        664 LMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMA 723
              |||||||||||::::|  |  :|     :  |  |||||      :||:||     | |:| :
Db       2700 LMQALPAGGAMVAVQAAEEEVRPLL---SAEVDIAAVNGPASLVISGAQNAVAAVADQLR 2756

Qy        724 ARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPG 783
              |  |  |           |  |||||||||| |||:   ||   ::  ||:|  :|||::|     |:   |
Db       2757 ADGRRVHQLAVSHAFHSPLMDPMIDEFAAVAAGIAIGRPTIGVISNVTGQLAGDDFGSAA 2816

Qy        784 YWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRD 843
              || || |: |||||  |:      |||     |:||||    |:   :     :||       ::: |   |
Db       2817 YWRRHIRQAVRFADSVRFAQAAGGSRFLEVGPSGGLVASIEESLPDVAVTTMSALRKDRP 2876

Qy        844 EPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI--DTKADDAA---- 897
              ||||:   |:      :    |   : |    :     | ||||  :||  |:|:   |   | |||
Db       2877 EPATLTNAVAQGFVTGMDLDWRAVV-GEAQFVELPTYAFQRRRFWLSGDGVAADAAGLGL 2935
```

```
Qy         898 --------------------------------------RGDRRAPGAGH---------DE 910
                                                       |    ||||         ||
Db        2936 AASEHALLGAVIDLPASGGVVLTGRLSPSVQGWLADHSVAGVTIFPGAGFVELAIRAGDE 2995

Qy         911 V-------------------------------EEG----------------------GAV 917
               |                                | |                      ||:
Db        2996 VGCGVVDESTLAAPLVLPASGSVAVQVVVNGPDESGVRGVSVYSRGDVGTGWVLHAEGAL 3055

Qy         918 RGG--------------------------------------------------------- 920
               | |
Db        3056 RAGSAEPTADLAMWPPAGAVPVEVADGYQQLAERGYGYGPAFRGLTAMWRRGDEVFAEVA 3115

Qy         921 ------------------------------------------------------------ 920

Db        3116 LPADAGVSVTGFGVHPVLLDAALHAVVLSAESAERGQGSVLVPFSWQGVSLHAAGASAVR 3175

Qy         921 -----------------------------------------------DR----------- 922
                                                              ||
Db        3176 ARIAPVGPSAVSIELADGLGLPVLSVASMLARPVTDQQLRAAVSSSGPDRLFEVTWSPQP 3235

Qy         923 ------------------------------------------------------------ 922

Db        3236 SAAVEPLPVCAWGTTEDSAAVVFESVPLAGDVVAGVYAATSSVLDVLQSWLTRDGAGVLV 3295

Qy         923 --------------------------RSARLDHP-------------------------P 931
                                         |||: :||                         |
Db        3296 VMTRGAVALPGEDVTDLAGAAVWGLVRSAQTEHPGRIVLVDSDAPLDDSALAAVVTTGEP 3355

Qy         932 PESGRREKVEAA---------------GDRPFRLEIDEPGVLDHLVLR-VTERRAP-GLG 974
                   || :|  |               |||:|| :   |  ::|  |  : :   || | |
Db        3356 QVLWRRGEVYTARVHGSRAVGGLLVPPSDRPWRLAMSTAGTFENLRLELIPDADAPLGPG 3415

Qy         975 EVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQP 1034
               :| :||  |    :| || :|||: ||        |   ::|  |  | ::         ||
Db        3416 QVRVAVSAIAANFRDVMIALGLYPD-------PDAVMGVEACGVVIETSLNKGSFAVGDR 3468

Qy        1035 VIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLI 1094
               |: |       |  :|   |::  |   |   ||   | : || |||  :|  : |:||||
Db        3469 VMGLFPEGTGTVASTDQRLLVKVPAGWSHTAAATTSVVFATAHYALVDLAAARSGQRVLI 3528

Qy        1095 HAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGV--RYVSDSRSDRFVADVRAW 1152
               || |||||:||||  |:|:| || |||  :   |   |  ::|     ::|||||  |   ||
Db        3529 HAGTGGVGMAAVQLARHLGLEVFATA-SKGKWDTLRAMGFDDDHISDSRSLEFEDKFRAA 3587

Qy        1153 TGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR-PFLRNLSFSLV 1211
               ||| | ||||:||:|| :| |  |:  | |:|:|| |        :  : | :|  :| |
Db        3588 TGGRGFDVVLDSLAGEFVDASLRLVAPGGVFLEMGKTDIRDPGVIAQQYPGVRYRAFDLF 3647

Qy        1212 DLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLV- 1270
               :          | |:  :| ||  |    ||  | |: |   :       | |  ::||:| ||:|
Db        3648 E------PGPDRIAQILAELATLFGDGVLRPLPVTTFDVRCAPAALRYLSQARHTGKVVM 3701

Qy        1271 ------------LTLGDPEVQIRIPTHAGA-------------GP-STGDRDLLDRLASA 1304
                           :||      :  |  |             || : |  :|:  ||:|
Db        3702 LMPGSWAAGTVLITGGTGMAGSAVARHVVARHGVRNLVLVSRRGPDAPGAAELVAELAAA 3761

Qy        1305 APAARAAALEAFLRTQVSQVLRTPEIK---------VGA--EALFTRLGMDSLMAVELRN 1353
                     :  | :|  |  :::|:    ::          ||  :|:    |  |  :  | ||:
Db        3762 GAQVQVVACDAADRAALAKVIADIPVQHPLSGVIHTAGALDDAVVMSLTPDRVDVV-LRS 3820

Qy        1354 RIEASLKLKLSTTFLSTSPNIAL----------------LAQNLLDALA 1386
               :::|:  |   |  |  |   :                 | :|||||
Db        3821 KVDAAWHLHELTRDLDVSAFVMFSSMAGLVGSSGQANYAAANSFLDALA 3869
```

Search completed: October 2, 2003, 17:31:44
Job time : 167.882 secs

```
OM protein - protein search, using sw model

Run on:         October  2, 2003, 17:27:29 ; Search time 26.2576 Seconds
                                             (without alignments)
                                             2544.975 Million cell updates/sec

Title:          US-10-014-717-2
Perfect score:  7210
Sequence:       1 VADRPIERAAEDPIAIVGAS..........GVQNDFVSSGADQDWEIIAL 1421

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:      127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries.

Database :       SwissProt_41:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2107.5	29.2	3567	1	ERY2_SACER	Q03132 saccharopol
2	1925.5	26.7	3491	1	ERY1_SACER	Q03131 saccharopol
3	1889	26.2	3519	1	OL56_STRAT	Q07017 streptomyce
4	1875	26.0	3172	1	ERY3_SACER	Q03133 saccharopol
5	1852.5	25.7	2110	1	MCAS_MYCBO	Q02251 mycobacteri
6	1592.5	22.1	1538	1	PPSB_MYCTU	Q10978 mycobacteri
7	1583	22.0	1876	1	PPSA_MYCTU	Q10977 mycobacteri
8	1180	16.4	1986	1	WA_EMENI	Q03149 emericella
9	1177	16.3	1774	1	MSAS_PENPA	P22367 penicillium
10	1173.5	16.3	2109	1	PKS1_ASPPA	Q12053 aspergillus
11	1100	15.3	2505	1	FAS_RAT	P12785 rattus norv
12	1083	15.0	2181	1	STCA_EMENI	Q12397 emericella
13	1039	14.4	2504	1	FAS_HUMAN	P49327 homo sapien
14	1018	14.1	4447	1	PKSK_BACSU	P40803 bacillus su
15	997.5	13.8	2511	1	FAS_CHICK	P12276 gallus gall
16	940	13.0	4427	1	PKSL_BACSU	Q05470 bacillus su
17	882.5	12.2	4273	1	PKSM_BACSU	P40872 bacillus su
18	612.5	8.5	352	1	FAS_ANSAN	P36189 anser anser
19	392.5	5.4	317	1	FABD_BACSU	P71019 bacillus su
20	363	5.0	401	1	NOE4_RHIME	P06231 rhizobium m
21	354.5	4.9	402	1	NODE_RHIME	P06230 rhizobium m
22	344	4.8	308	1	FABD_SALTY	O85140 salmonella
23	336	4.7	403	1	NODE_RHILV	P04683 rhizobium l
24	332	4.6	308	1	FABD_ECOLI	P25715 escherichia
25	329.5	4.6	467	1	KASA_STRCO	Q02059 streptomyce
26	328	4.5	312	1	FABD_HAEIN	P43712 haemophilus
27	325.5	4.5	402	1	NODE_RHIS3	P72331 rhizobium s
28	324.5	4.5	401	1	NODE_RHILT	P04684 rhizobium l
29	321.5	4.5	426	1	KAS1_STRGA	P16538 streptomyce
30	316.5	4.4	300	1	VAT1_HUMAN	Q99536 homo sapien
31	315	4.4	412	1	FABF_ECOLI	P39435 escherichia
32	308	4.3	414	1	FABF_VIBHA	P55338 vibrio harv
33	304.5	4.2	293	1	FABD_SYNY3	P73242 synechocyst
34	303.5	4.2	413	1	FABF_VIBCH	Q9kqh9 vibrio chol
35	303	4.2	838	1	FAS_MOUSE	P19096 mus musculu
36	302	4.2	420	1	KAS1_STRCM	P41175 streptomyce
37	298.5	4.1	327	1	QOR_ECOLI	P28304 escherichia
38	298	4.1	421	1	KAS1_STRVN	P16540 streptomyce
39	295.5	4.1	327	1	QOR_SALTY	P40783 salmonella
40	293	4.1	421	1	FABF_RHIME	P56902 rhizobium m
41	289.5	4.0	422	1	KAS1_STRHA	Q05356 streptomyce
42	285.5	4.0	416	1	FABF_SYNY3	P73283 synechocyst
43	284.5	3.9	423	1	KAS1_STRCO	P23155 streptomyce
44	282	3.9	379	1	VAT1_TORCA	P19333 torpedo cal
45	278	3.9	325	1	QOR_PSEAE	P43903 pseudomonas

ALIGNMENTS

```
RESULT 1
ERY2_SACER
ID   ERY2_SACER     STANDARD;      PRT;  3567 AA.
AC   Q03132; Q54096;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE   deoxyerythronolide B synthase II) (DEBS 2).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: Contains 2 acyl carrier domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   PIR; S23070; S23070.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; ADH_zinc_N; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 3.
FT   DOMAIN     1485   3567       MODULE 4.
FT   DOMAIN       27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                NON-FUNCTIONAL).
FT   DOMAIN     1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN     3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    202    202       THIOESTER BOND.
FT   ACT_SITE    651    651       ACYL-ESTER INTERMEDIATE.
FT   BINDING    1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2105   2105       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    2961   2978       NADP (ER).
FT   NP_BIND    3142   3157       NADP (KR).
FT   BINDING    3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    438    438       R -> A (IN REF. 2).
FT   CONFLICT    480    480       T -> S (IN REF. 2).
FT   CONFLICT   1241   1241       L -> F (IN REF. 2).
FT   CONFLICT   2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match             29.2%;  Score 2107.5;  DB 1;  Length 3567;
  Best Local Similarity   32.5%;  Pred. No. 4.5e-103;
  Matches  610;  Conservative  179;  Mismatches  530;  Indels  559;  Gaps   45;

Qy          2 ADRPIERAAE-DPIAIVGASCRLPGGVIDLSGFWTLL-EGSRDTVGRVPAER-WDAAAWF 58
              | | :   | | :||||||  || |||:        | :| ||:  | |   ||:| ||    :
Db       1478 AVRVVGAADESEPIAIVGIGCRFPGGIGSPEQLWRVLAEGANLTTG-FPADRGWDIGRLY 1536

Qy         59 DPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIA 118
              |||| || : | :  ||:| | ||  ||||:||||| |||   ||:||  |||:| | |
Db       1537 HPDPDNPGTSYVDKGGFLTDAADFDPGFFGITPREALAMDPQQRLMLETAWEAVERAGID 1596

Qy        119 PSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPC 178
              | || ||:||||:|:    |   |   :  :| : |||   ||  | :|||:|  |   ||
Db       1597 PDALRGTDTGVFVGMNGQSYMQLL--AGEAERVDGYQGLGNSASVLSGRIAYTFGWEGPA 1654

Qy        179 VAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKA 238
              : |||| ||||| :||| |:|| |||| ||||||::|  |  |   |  | || ||||||
Db       1655 LTVDTACSSSLVGIHLAMQALRRGECSLALAGGVTVMSDPYTFVDFSTQRGLASDGRCKA 1714

Qy        239 FSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEI 298
              ||| ||||    || |  :||:  || |||:| ::|||:||||:| ||||:||  |||  |||
Db       1715 FSARADGFALSEGVAALVLEPLSRARANGHQVLAVLRGSAVNQDGASNGLAAPNGPSQER 1774

Qy        299 VLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNL 358
              |:::|||  :|   |: |  |||||||| ||||||  || | ||   ||    || :||||||:
Db       1775 VIRQALAASGVPAADVDVVEAHGTGTELGDPIEAGALIATYGQDRD--RPLRLGSVKTNI 1832

Qy        359 GHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPR 418
              || : |:|  |::|||:::|| :|  |||  |:| | |       : | |    |||    ||
Db       1833 GHTQAAAGAAGVIKVVLAMRHGMLPRSLHADELSPHIDWESGAVEVLREEVPWPAGERPR 1892

Qy        419 RAGVSSFGMSGTNAHVVLEEAP-----AATCTPPAPERPAELLVLSARTASALDAQAARL 473
              |||||||:|||||||::||||       | |     | |        ||| |: : : ||| |
Db       1893 RAGVSSFGVSGTNAHVIVEEAPAEQEAARTERGPLP------FVLSGRSEAVVAAQARAL 1946

Qy        474 RDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSI 533
               :||   |    | | |::|||  |:   :  | ||      | |:  | ||| |:|:  |  || :
Db       1947 AEHLRDTPELGLTDAAWTLATGRARFDVRAAVLGDDRAGVCAELDALAEGRPSADAVAPV 2006

Qy        534 ADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEP 593
               : |  |   :| ||||| :|| | |  :       ||  |:    |       :     |   | :|: :
Db       2007 TSAPR-KPVLVFPGQGAQWVGMARDLLESSEVFAESMSRCAEALSPHTDWKLLDVVRGDG 2065

Qy        594 ASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLED 653
                :|     || || |:   :|| |||: || |   | ||| ||: || ||| |||
Db       2066 GPDPHERVD---VLQPVLFSIMVSLAELWRAHGVTPAAVVGHSQGEIAAAHVAGALSLEA 2122

Qy        654 AVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQ 713
              |   :|| | ::::: |    | |||: |    ::   :|     |::|||| |    |:||
Db       2123 AAKVVALRSQVLRELDDQGGMVSVGASRDELETVLARWDGRVAVAAVNGPGTSVVAGPTA 2182

Qy        714 PVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGK 773
               :     ||  : : : :  | :| ||| :| : :          :::  |  | |   | ::|:
Db       2183 ELDEFFAEAEAREMKPRRIAVRYASHSPEVARIEDRLAAELGTITAVRGSVPLHSTVTGE 2242

Qy        774 ACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLL---------GLV 823
                   || |: |   |  |    |: |     |  |||||  |    ||          |
Db       2243 VIDTSAMDASYWYRNLRRPVLFEQAVRGLVEQGFDTFVEVSPHPVLLMAVEETAEHAGAE 2302

Qy        824 PACMPDARPALLASSRAGRDEPATVLEAL--GGLWAVGGLVSWAGLFP--SGGRRVPLPT 879
               |:|          :  |  :  |   |   |   |  :   ||    |  |  |   :|||   |||
Db       2303 VTCVP--------TLRREQSGPHEFLRNLLRAHVHGVG-----ADLRPAVAGGRPAELPT 2349

Qy        880 YPWQRERYW--IDTKADDAARGDRRA---------------------------------- 903
              ||:: :|:|        || :| | | |
Db       2350 YPFEHQRFWPRPHRPADVSALGVRGAEHPLLLAAVDVPGHGGAVFTGRLSTDEQPWLAEH 2409

Qy        904 --------PG---------AGHD------------------------------------- 909
                      ||         | | |
Db       2410 VVGGRTLVPGSVLVDLALAAGEDVGLPVLEELVLQRPLVLAGAGALLRMSVGAPDESGRR 2469

Qy        910 ------------------------------------------------------------ 909

Db       2470 TIDVHAAEDVADLADAQWSQHATGTLAQGVAAGPRDTEQWPPEDAVRIPLDDHYDGLAEQ 2529

Qy        910 ------------------------EV-----EEG------------------GAV------- 917
                                      ||     |||                  ||:
Db       2530 GYEYGPSFQALRAAWRKDDSVYAEVSIAADEEGYAFHPVLLDAVAQTLSLGALGEPGGGK 2589

Qy        918 -------------------------RGGDRRSARL-----------------------DHP 930
                                       | |  : |:                       : |
Db       2590 LPFAWNTVTLHASGATSVRVVATPAGADAMALRVTDPAGHLVATVDSLVVRSTGEKWEQP 2649

Qy        931 PPESGRRE------------------------------------------KVEAAGDRP- 947
              |   |  |                                           : |  || |
Db       2650 EPRGGEGELHALDWGRLAEPGSTGRVVAADASDLDAVLRSGEPEPDAVLVRYEPEGDDPR 2709

Qy        948 ------------------------------------------------------------ 947

Db       2710 AAARHGVLWAAALVRRWLEQEELPGATLVIATSGAVTVSDDDSVPEPGAAAMWGVIRCAQ 2769

Qy        948 -------FRLEID-EPGVL---------------DHLVLRV-------------TERRAPG 972
                     |: |  |||:|               |  | |:               |:|  ||
Db       2770 AESPDRFVLLDTDAEPGMLPAVPDNPQLALRGDDVFVPRLSPLAPSALTLPAGTQRLVPG 2829

Qy        973 -------------------LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012
                                 ||| :  |  | |::| || ||||| |          :|
Db       2830 DGAIDSVAFEPAPDVEQPLRAGEVRVDVRATGVNFRDVLLALGMYPQKAD--------MG 2881
```

```
Qy       1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072
                | ||  : |||  |:      |  |: |   ||||     |    |:     |     |   :|||:|:|
Db       2882 TEAAGVVTAVGPDVDAFAPGDRVLGLFQGAFAPIAVTDHRLLARVPDGWSDADAAAVPIA 2941

Qy       1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132
                | ||`|||   :| |:  |: ||||| ||||:|||   |:    |||| |||| |  |      |  :|
Db       2942 YTTAHYALHDLAGLRAGQSVLIHAAAGGVGMAAVALARRAGAEVLATAG-PAKHGTLRAL 3000

Qy       1133 GV--RYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRD 1190
              |:    ::: ||    |       |    ||| ||||||||:|||:|:|  :||     |  |||:|| |
Db       3001 GLDDEHIASSRETGFARKFRERTGGRGVDVVLNSLTGELLDESADLLAEDGVFVEMGKTD 3060

Qy       1191 CYADNQLGLGLRPFLRNLSFSLVDLRGMMLERPA----------RVRALLEELLGLIAAGVF 1240
                                   ||:       | ||    | |              |:   :|  |::||: ||
Db       3061 ------------LRDAG----DFRG----RYAPFDLGEAGDDRLGEILREVVGLLGAGEL 3100

Qy       1241 TPPPIATLPIARVADAFRSMAQAQHLGKLVLTLG---DPEVQIRIPTHAGAGPSTGDRDL 1297
                       |::    :         |  :  |:: :|:|||||        ||:    :  |                   ||
Db       3101 DRLPVSAWELGSAPAALQHMSRGRHVGKLVLTQPAPVDPDGTVLI---------TGGTGT 3151

Qy       1298 LDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEA 1357
                | ||        ||      |    :|     :                         | :|  |  |:   :  |||   ||
Db       3152 LGRL-----LARHLVTEHGVRHLL---------------LVSRRGADAPGSDELRAEIE- 3190

Qy       1358 SLKLKLSTTFLSTSPNIALLAQNLLDALATAL----------------SLERVAAEN-LR 1400
                 |               |:    ||  |    |||  |      |                              |::    |  |    ||
Db       3191 DLGASAEIAACDTADRDALSA--LLDGLPRPLTGVVHAAGVLADGLVTSIDEPAVEQVLR 3248

Qy       1401 AGVQNDFVSSGADQDWEI 1418
                | |            |  |  :
Db       3249 AKV---------DAAWNL 3257
```

```
RESULT 2
ERY1_SACER
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: Contains 3 acyl carrier domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   PIR; T43231; T43231.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1972       MODULE 1.
FT   DOMAIN     1979   3491       MODULE 2.
FT   DOMAIN        1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN      503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1888   1958       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN     3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3334   3404       ACYL CARRIER (ACP) 3.
FT   ACT_SITE    145    145       ACYL-ESTER INTERMEDIATE.
FT   BINDING     447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   1128   1128       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    1614   1660       NADP.
FT   BINDING    1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2598   2598       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    3058   3104       NADP.
FT   BINDING    3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match             26.7%;  Score 1925.5;  DB 1;  Length 3491;
  Best Local Similarity   35.6%;  Pred. No. 1.7e-93;
  Matches  547;  Conservative  184;  Mismatches  601;  Indels  203;  Gaps   39;

Qy          4 RPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDP 62
                | |  |   :|:|:|  :||||||        || ||    ||  |   :|  :|  ||   :  | |||
Db        497 RTNEAAPGEPVAVVAMACRLPGGVSTPEEFWELLSEGRDAVAGLPTDRGWDLDSLFHPDP 556

Qy         63 DAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSAL 122
                      |           ||::    || :|||:||||||  :||   ||:||:  || || |  |  |::|
Db        557 TRSGTAHQRGGGFLTEATAFDPAFFGMSPREALAVDPQQRLMLELSWEVLERAGIPPTSL 616
```

```
Qy    123 VGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVD 182
Db    617 QASPTGVFVGLIPQEYGPRL--AEGGEGVEGYLMTGTTTSVASGRIAYTLGLEGPAISVD 674

Qy    183 TAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAE 242
Db    675 TACSSSLVAVHLACQSLRRGESSLAMAGGVTVMPTPGMLVDFSRMNSLAPDGRCKAFSAG 734

Qy    243 ADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKR 302
Db    735 ANGFGMAEGAGMLLLERLSDARRNGHPVLAVLRGTAVNSDGASNGLSAPNGRAQVRVIQQ 794

Qy    303 ALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPE 362
Db    795 ALAESGLGPADIDAVEAHGTGTRLGDPIEARALFEAY--GRDREQPLHLGSVKSNLGHTQ 852

Qy    363 YASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGV 422
Db    853 AAAGVAGVIKMVLAMRAGTLPRTLHASERSKEIDWSSGAISLLDEPEPWPAGARPRRAGV 912

Qy    423 SSFGMSGTNAHVVLEEAP---------AATCTPPAPERPAELLVLSARTASALDAQAARL 473
Db    913 SSFGISGTNAHAIIEEAPQVVEGERVEAGDVVAP--------WVLSASSAEGLRAQAARL 964

Qy    474 RDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSI 533
Db    965 AAHLREHPGQDPRDIAYSLATGRAALPHRAAFAPVDESAALRVLDGLATGNADGAAVGT- 1023

Qy    534 ADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEP 593
Db   1024 -SRAQQRAVFVFPGQGWQWAGMAVDLLDTSPVFAAALRECADALEPHLDFEVIPFLRAEA 1082

Qy    594 A--SVDAAL-LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFS 650
Db   1083 ARREQDAALSTERVDVVQPVMFAVMVSLASMWRAHGVEPAAVIGHSQGEIAAACVAGALS 1142

Qy    651 LEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAVAPHAASVSIAAVNAPDQVVIAG 710
Db   1143 LDDAARVVALRSRVIATMPGNKGMASIAAPAGEVRARIGDR---VEIAAVNGPRSVVVAG 1199

Qy    711 AGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIV-LVSN 769
Db   1200 DSDELDRLVASCTTECIRAKRLAVDYASHSSHVETIRDALHAELGEDFHPLPGFVPFFST 1259

Qy    770 LSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPD 829
Db   1260 VTGRWTQPDELDAGYWYRNLRRTVRFADAVRALAEQGYRTFLEVSAHPILTAAIEEIGDG 1319

Qy    830 ARPALLA--SSRAGRDEPATVLEALGGLWAVGGLVSWAGL-FPSGGRRVPLPTYPWQRER 886
Db   1320 SGADLSAIHSLRRGDGSLADFGEALSRAFAAGVAVDWESVHLGTGARRVPLPTYPFQRER 1379

Qy    887 YWIDTKADDAARGDRRAPGAHGDEVEEGGAVR-------GGDRRSARLD----------- 928
Db   1380 VWLEPK--PVARRS--------TEVDEVSALRYRIEWRPTGAGEPARLDGTWLVAKYAGT 1429

Qy    929 HPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPGLGEVE-----IAVDAA 983
Db   1430 ADETSTAAREALESAGARVRELVVDARCGRDELA-----ERLRSVGEVAGVLSLLAVDEA 1484

Qy    984 GLSFNDVQLALGMVPDDLP----------GKPNPPLLLGGECAGRIVAVGEGVNGLV--V 1031
Db   1485 --EPEEAPLALASLADTLSLVQAMVSAELGCPLWTVTESAVATGPFERVRNAAHGALWGV 1542

Qy   1032 GQPVIALSAGAFATHVTTSAALVLPRPQALSAIE---------AAAMPVAYLTAWYALDR 1082
Db   1543 GRVIALENPAVWGGLVDVPAGSVAELARHLAAVVSGGAGEDQLALRADGVYGRRWVRAAA 1602

Qy   1083 IA---RLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHA---------TAGTPEKRAYLE 1130
Db   1603 PATDDEWKPTGTVLVTGGTGGVGGQIARWLARRGAP-HLLLVSRSGPDADGAGELVAELE 1661

Qy   1131 SLGVRY------VSDSRSDRFVADVRAWTGGEGVDVVL---------------NSLSGEL 1169
Db   1662 ALGARTTVAACDVTDRES------VRELLGGIGDDVPLSAVFHAAATLDDGTVDTLTGER 1715

Qy   1170 IDKSFNLLRSHGRFVELGKRDCY-ADNQLGLRPFLRNLSFSL--------------VDLR 1214
Db   1716 IERA-------SRAKVLGARNLHELTRELDLTAFVLFSSFASAFGAPGLGGYAPGNAYLD 1768

Qy   1215 GMMLER-----PARVRA---------LLEELLGLIA-----AGVFTPPP----------- 1244
Db   1769 GLAQQRRSDGLPATAVAWGTWAGSGMAE---GAVADRFRRHGVIEMPPETACRALQNALD 1825

Qy   1245 ---IATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRL 1301
Db   1826 RAEVCPIVIDVRWDRFLLAYTAQRPTRLFDEIDDAR---RAAPQAPAEPRVG------AL 1876

Qy   1302 ASAAPAARAAALEAFLRTQVSQVL-RTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLK 1360
Db   1877 ASLPAPEREEALFELVRSHAAAVLGHASAERVPADQAFAELGVDSLSALELRNRLGAATG 1936

Qy   1361 LKLSTTFLSTSPNIALLAQNLLDALATALSLERVA 1395
Db   1937 VRLPTTTVFDHPDVRTLAAHLAAELGGATGAEQAA 1971
```

RESULT 3
OL56_STRAT

```
ID   OL56_STRAT     STANDARD;      PRT;  3519 AA.
AC   Q07017;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Oleandomycin polyketide synthase, modules 5 and 6.
GN   ORFB.
OS   Streptomyces antibioticus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150470; PubMed=8107683;
RA   Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT   "Characterisation of a Streptomyces antibioticus gene encoding a type
RT   I polyketide synthase which has an unusual coding sequence.";
RL   Mol. Gen. Genet. 242:358-362(1994).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC       LACTONE RING.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- SIMILARITY: Contains 2 acyl carrier domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L09654; AAA19695.1; -.
DR   PIR; S43048; S43048.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
```

```
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN        1      ?       MODULE 5.
FT   DOMAIN        ?   3519       MODULE 6.
FT   DOMAIN       32    501       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      569    890       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1200   1382       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1487   1561       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1686   2156       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2220   2541       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2856   3038       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3141   3215       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3270   3519       THIOESTERASE.
FT   ACT_SITE    210    210       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    660    660       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    1203   1249       NADP.
FT   BINDING    1524   1524       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1859   1859       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2311   2311       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    2859   2905       NADP.
FT   BINDING    3178   3178       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3519 AA;  368561 MW;  41AE78AAAEB61F86 CRC64;

  Query Match             26.2%;  Score 1889;  DB 1;  Length 3519;
  Best Local Similarity   34.0%;  Pred. No. 1.4e-91;
  Matches  541;  Conservative  191;  Mismatches  595;  Indels  264;  Gaps   44;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGK 67

Db       1684 ASREAIAITAMSCRFGGGIDSPEDLWRFLAEGRDAVAGLPEDRGWDLDALYHPDPENPGT 1743

Qy         68 TPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTET 127

Db       1744 TYVREGAFRYDAAQFDAGFFGISPREALAMDPQQRLLLETSWELFERADIDPYTVRGTAT 1803

Qy        128 GVFIGIGPSEY---EAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184

Db       1804 GIFIGAGHQGYGPDPKRAPESVAGYLL-----TGTASAVLSGRISYTFGLEGPAVTVDTA 1858

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244

Db       1859 CSSSLVALHLAVQALRRGECSLAIAGGVAVMSTPDAFVEFSRQQGMARDGRCKAFAAAAD 1918

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304

Db       1919 GMGWGEGVSLLLLERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAAL 1978

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364

Db       1979 ADAGLAPADVDVVEAHGTGTRLGDPIEAQALLATYGQGRAGGRPVWLGSVKSNIGHTQAA 2038

Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPW-PDWNTPRRAGVS 423

Db       2039 AGVAGVMKMVLALGRGVVPKTLHVDEPSPHVDWSAGAVELLTEERPWEPEAERLRRAGIS 2098

Qy        424 SFGMSGTNAHVVLEEAPAATCTPPAPERPAELL---------VLSARTASALDAQAARLR 474

Db       2099 AFGVSGTNAHVIVEEAPA----EPEPEPGTRVVAAGDLVVPWVVSGRDVGALREQAARLA 2154

Qy        475 DHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIA 534

Db       2155 AHVSSTGAGVV-DVGWSLVATRSVFEHRAVMVGTDLDSMAGSLAGFAAGGVVPGVVSGVA 2213

Qy        535 DSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPA 594

Db       2214 PAEGRRVVFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLQGR-- 2271

Qy        595 SVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDA 654

Db       2272 --DATVLGRVDVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADG 2329

Qy        655 VFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQP 714

Db       2330 ARVVVLRSRAIARIAGGGGMVSVSLPAGRVRTMLDTYGGRVSVAAVNGPSSTVVSGDVQA 2389

Qy        715 VHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKA 774

Db       2390 LDELLAGCEREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQDSSVPFFSTVTADW 2449

Qy        775 CTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMP--DARP 832

Db       2450 LDTTALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQNA 2509

Qy        833 ALLASSRAGRDEPA--TVLEALGGLWAVGGLVSWA----GLFPSGGRRVPLPTYPWQRER 886

Db       2510 AVLGSLR--RDEGGLDRLLTSLAEAFVQGVPVDWTHAFEGVTP---RTVDLPTYPFQRQR 2564

Qy        887 YWIDTKADDAARG-DRRAPGAGHDEVEEGGAVR-----GGDRRSARLDHPPPESGRREKV 940

Db       2565 FWLDGSPASSANGVDGEADAMIWDAVEREDSVAVAEELGIDAEALHTVLPALSSWRRRRV 2624

Qy        941 E---------AAGDRPPFRLEIDE---------------PGVLDHLVLRVTERRAPGLGE 975

Db       2625 EHRRLQDWRYRVEWKPFPAALDEVLGGGWLFVVPRGLADDGVVARVVAAVTARG----GE 2680

Qy        976 VEI--------------AVDAAGLSFNDVQLALGMVP----DDLPGKPNPPLLLGGECA 1016

Db       2681 VSVVELDPTRPDRRAYAEAVAGRGVS--------GVVSFLSWDDRRHSEH-PVVPAGLAA 2731

Qy       1017 GRIVA--------VGEGVN-GLVVGQPVIA--LSAGAFATHVTTSA-----ALVLPRPQ- 1059

Db       2732 SLVLAQALVDLGRVGEGPRLWLVTRDAVVAGPSDAGAVIDPVQAQVWGFGRVLGLEHPEL 2791

Qy       1060 -------------------------ALSAIEAAA-----MPVAYLTAWYALDRIARL---- 1086

Db       2792 WGGLIDLPVEAPEPGSTCDHTYADLLATVVASAGFEDQVAVRGSGVW--VRRLVRAVVDG 2849

Qy       1087 -----QPGERVLIHAATGGVGLAAVQWAQHVGAEVH---------ATAGTPEKRAYLESL 1132

Db       2850 GGGGWRPRGTVLVTGGLGGLGAHTARWLVGGGAD-HVVLVSRRGGSAPGAGDLVRELEGL 2908

Qy       1133 GVRYVSDSRSDRFVAD---VRAWTG--GEGVDVVLNSLS--------------------- 1166

Db       2909 GGARVSVRACD--VADRVALRALLSDLGEPVTAVFHAAGVPQSTPLAEISVQEAADVMAA 2966

Qy       1167 --------GELID----KSFNLLRSHGRFVELGKRDCYA--------------------- 1193

Db       2967 KVAGAVNLGELVDPCGLEAFVLFSSNAGVWGSGGQAVYAAANAFLDALAVRRRGVGLPAT 3026

Qy       1194 ---------DNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPP 1244

Db       3027 SVAWGMWAGEGMASVGGAARELS-----RRGVRAMDPERAVAVMADAVGRGEAFV----A 3077

Qy       1245 IATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRI------PTHAGAGPSTGDRDLL 1298

Db       3078 VADVDWERFVTGFASARPRPLISDL------PEVRTALRNQEQEQLHAPV-PEDRSAQLL 3130

Qy       1299 DRLASAAPAARAAALEAFLRTQVSQVLRTPEIK-VGAEALFTRLGMDSLMAVELRNRIEA 1357

Db       3131 RRLSMLSPAGREAELVKLVRTEAAAVLGHGSAQDVPAERAFKELGFDSLTAVQLRNRLAA 3190

Qy       1358 SLKLKLSTTFLSTSPNIALLAQNLLDALATA 1388

Db       3191 ATGTRLPASAVFDHPHAAALARWLLAGMRHA 3221

RESULT 4
ERY3_SACER
ID   ERY3_SACER     STANDARD;      PRT;  3172 AA.
AC   Q03133; Q54097; Q99270;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE   deoxyerythronolide B synthase III) (DEBS 3).
GN   ERYA.
```

```
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=91043075; PubMed=2234082;
RA   Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT   "An unusually large multifunctional polypeptide in the erythromycin-
RT   producing polyketide synthase of Saccharopolyspora erythraea.";
RL   Nature 348:176-178(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: Contains 2 acyl carrier domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   PIR; S13595; S13595.
DR   PIR; S22012; S22012.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 5.
FT   DOMAIN     1485   3172       MODULE 6.
FT   DOMAIN       37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2926   3172       THIOESTERASE.
FT   ACT_SITE    199    199       THIOESTER BOND.
FT   ACT_SITE    643    643       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    1118   1164       NADP.
FT   BINDING    1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2112   2112       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    2557   2605       NADP.
FT   BINDING    2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    231    231       S -> SA (IN REF. 2).
FT   CONFLICT    240    240       MISSING (IN REF. 2).
FT   CONFLICT    289    289       A -> R (IN REF. 2).
FT   CONFLICT    493    493       P -> R (IN REF. 2).
FT   CONFLICT    493    517       PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT                                RWPPAAAPWEQ (IN REF. 1).
FT   CONFLICT    510    510       A -> R (IN REF. 2).
FT   CONFLICT    513    513       M -> W (IN REF. 2).
FT   CONFLICT    525    525       E -> D (IN REF. 2).
FT   CONFLICT    536    536       R -> G (IN REF. 2).
FT   CONFLICT    547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT    553    553       R -> G (IN REF. 2).
FT   CONFLICT    673    673       R -> A (IN REF. 2).
FT   CONFLICT    716    716       MISSING (IN REF. 2).
FT   CONFLICT    734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT    896    896       R -> RQR (IN REF. 2).
FT   CONFLICT    896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT    988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT   1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT   1124   1126       MISSING (IN REF. 1).
FT   CONFLICT   1132   1132       L -> V (IN REF. 2).
FT   CONFLICT   1192   1192       A -> R (IN REF. 2).
FT   CONFLICT   1194   1194       MISSING (IN REF. 2).
FT   CONFLICT   1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT   1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT   1485   1485       MISSING (IN REF. 2).
FT   CONFLICT   1518   1518       G -> R (IN REF. 2).
FT   CONFLICT   1601   1601       V -> L (IN REF. 2).
FT   CONFLICT   1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT   1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT   1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT   1762   1762       T -> S (IN REF. 2).
FT   CONFLICT   2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT   2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT   2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT   2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT   2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT   2596   2596       A -> G (IN REF. 2).
FT   CONFLICT   2609   2609       P -> A (IN REF. 2).
FT   CONFLICT   2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT   2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match             26.0%;  Score 1875;  DB 1;  Length 3172;
  Best Local Similarity   34.6%;  Pred. No. 6.9e-91;
  Matches  549;  Conservative  184;  Mismatches  602;  Indels  254;  Gaps   48;
```

```
Qy          4 RPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPD 63
Db         31 REVEHRAGEPIAIVGMACRFPGDVDSPESFWEFVSGGGDAIAEAPADR----GW-EPDPD 85

Qy         64 APGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALV 123
Db         86 A------RLGGMLAAAGDFDAGFFGISPREALAMDPQQRIMLEISWEALERAGHDPVSLR 139

Qy        124 GTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDT 183
Db        140 GSATGVFTGVGTVDYGPRPDE--APDEVLGYVGTGTASSVASGRVAYCLGLEGPAMTVDT 197

Qy        184 AYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEA 243
Db        198 ACSSGLTALHLAMESLRRDECGLALAGGVTVMSSPGAFTEFRSQGGLAADGRCKPFSKAA 257

Qy        244 DGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRA 303
Db        258 DGFGLAEGAGVLVLQRLSAARREGRPVLAVLAGSAVNQDGASNGLTAPSGPAQQRVIRRA 317

Qy        304 LADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEY 363
Db        318 LENAGVRAGDVDYVEAHGTGTRLGDPIEVHALLSTYGAERDPDDPLWIGSVKSNIGHTQA 377

Qy        364 ASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVS 423
Db        378 AAGVAGVMKAVLALRHGEMPRTLHFDEPSPQIEWDLGAVSVVSQARSWPAGERPRRAGVS 437

Qy        424 SFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQ 483
Db        438 SFGISGTNAHVIVEEAPEADEPEPAPDSGPVPLVLSGRDEQAMRAQAGRLADHLAPEPRN 497

Qy        484 CLGDVAFSLATTRSAMEHRLAVAATSRE---GLRAALDAAAQGQTSPGAVRSIADSSRGK 540
Db        498 SLRDTGFTLATRASAMEHRAVVVGDRDEALAGLRAVADRRIADRTATGQ----GPNSPRR 553

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVM-WAEPASVDAA 599
Db        554 VAMVFPGQGAQWQGMARDLLRESQVFADSIRDCERALAPHVDWSLTDLLSGARP------ 607

Qy        600 LLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVA 659
Db        608 -LDRVDVVQPALFAVMVSLAALWRSHGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVA 666

Qy        660 ARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAP-DQVVIAGAGQPVHAI 718
Db        667 VRSRVLRRLGGQGGMASFGLGTEQAAERIGRFAGALSIASVNGPRSVVVVAGESGPLDEL 726

Qy        719 AAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDE 778
Db        727 IAECEAEAHKARRIPVDYASHSPQVESLREELLTELAGISPVSADVALYSTTTGQPIDTA 786

Qy        779 VSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPA----- 833
Db        787 TMDTAYWYANLREQVRFQDATRQLAEAGFDAFVEVSPHPVLTVGIEATLDSALPADAGAC 846

Qy        834 LLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKA 893
Db        847 VVGTLRRDRGGLADFHTALGEAYAQGVEVDWSPAF-ADARPVELPVYPFQ--RYWLPIPT 903

Qy        894 DDAAR---GDRRAPGAGHDEVEEGGAVRGGDRRSARLDHP--PPE--SGRREKVEAAGDR 946
Db        904 GGRARDEDDDWRYQVVWREAEWESASLAG---RVLLVTGPGVPSELSDAIRSGLEQSGAT 960

Qy        947 PFRLEID---------EPGVLDHL----VLRVTERRA--PGLGEVEI--AVDAAG----- 984
Db        961 VLTCDVESRSTIGTALEAADTDALSTVGVAAVPHGEAVDPSLDALALVQALGAAGVEAPL 1020

Qy        985 --LSFNDVQLALGMVPDDLPGKPNPPLLGGECAGRIVAVGE-----GVNGLVVGQPVIA 1037
Db       1021 WVLTRNAVQVADGELVD--PAQ----AMVGG--LGRVVGIEQPGRWGGLVDLV------- 1065

Qy       1038 LSAGAFATHVTTSAALVL-PRPQALSAIEAAAMPVAYLTAWYALDRIARLQP-GERVLIH 1095
Db       1066 ---DADAASIRSLAAVLADPRGEEQVAIRADGIKVARLVPAPARARTHPLEPLAGTVLVT 1122

Qy       1096 AATGGVGLAAVQWAQHVGAEVHA---------TAGTPEKRAYLESLGVRY------VSDS 1140
Db       1123 GGTGGIGAHLARWLARSGAE-HLVLLGRRGADAPGASELREELTALGTGVTIAACDVADR 1181

Qy       1141 RSDRFVADVRAWTGGEGVDVVLN----SLSGELIDKSFNLLRSHGRFVELGKRDCYADNQ 1196
Db       1182 ARLEAVLAAEAAAEGRTVSAVMHAAGVSTSTPLDD------LTEAEFTEI------AD-- 1227

Qy       1197 LGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIA----AGVFTPPPIATLPIAR 1252
Db       1228 -------------VKVRGTV-----NLDELCPDLDAFVLFSSNAGVWGSPGLASYAAAN 1268

Qy       1253 V-ADAFRSMAQA-----------------------QHLGKLVLTLGDP-----EVQIRI 1282
Db       1269 AFLDGFARAARSEGAPVTSIAWGLWAGQNMAGDEGGEYLRSQGLRAMDPDRAVEELHITL 1328

Qy       1283 PTHAGAGPSTGDRD---------------LLDRLASA-----------APAARAAA---- 1312
Db       1329 -DHGQTSVSVVDMDRRRFVELFTAARHRPLFDEIAGARAEARQSEEGPALAQRLAALLCD 1387

Qy       1313 ------LEAFLRTQVSQVL-RTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLST 1365
Db       1388 GREREHLAHLIRAEVAAVLGHGDDAAIDRDRAFRDLGFDSMTAVDLRNRLAAVTGVREAA 1447

Qy       1366 TFLSTSPNIALLAQNLLDALATALSLERVAAE-------------------NLRAGVQN 1405
Db       1448 TVVFDHPTITRLADHYLERLVGAAEAEQAPALVREVPPKDADDPIAIVGMACRFPGGVHN 1507

Qy       1406 -----DFVSSG--------ADQDWEIIAL 1421
Db       1508 PGELWEFIVGGGDAVTEMPTDRGWDLDAL 1536
```

RESULT 5
MCAS_MYCBO
ID   MCAS_MYCBO   STANDARD;   PRT; 2110 AA.
AC   Q02251;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Mycocerosic acid synthase.
GN   MAS.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92406887; PubMed=1527058;
RA   Mathur M., Kolattukudy P.E.;
RT   "Molecular cloning and sequencing of the gene for mycocerosic acid
RT   synthase, a novel fatty acid elongating multifunctional enzyme, from
RT   Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL   J. Biol. Chem. 267:19388-19395(1992).
CC   -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC   METHYLAMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC   FORM MYCOCEROSYL LIPIDS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.
CC   -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC   ARRANGEMENT.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: Contains 1 acyl carrier domain.
CC   --

```
CC   -----------------------------------------------------------------------
DR   EMBL; M95808; AAA25369.1; -.
DR   PIR; B44110; B44110.
DR   HSSP; P73283; 1E5M.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00107; ADH_zinc_N; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT   DOMAIN        1    430       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      533    852       ACYL TRANSFERASE.
FT   DOMAIN        ?      ?       ENOYL REDUCTASE.
FT   DOMAIN        ?      ?       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2026   2096       ACYL CARRIER (ACP).
FT   ACT_SITE    177    177       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    623    623       ACYL TRANSFERASES (BY SIMILARITY).
FT   NP_BIND    1561   1578       NADP (ER).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2059   2059       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   2110 AA;  225577 MW;  CC658215D7155300 CRC64;

  Query Match             25.7%;  Score 1852.5;  DB 1;  Length 2110;
  Best Local Similarity   29.5%;  Pred. No. 6.5e-90;
  Matches  524;  Conservative  176;  Mismatches  519;  Indels  555;  Gaps   30;

Qy         13 PIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKTPVTR 72
              |:|::|  |||||:       |  |     | |   :| :||||  ::||:|  ||::
Db          7 PVAVIGMGCRLPGGINSPDKLWESLLRGDDLVTEIPPDRWDADDYYDPEPGVPGRSVSRW 66

Qy         73 ASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGVFIG 132
                || ||| ||| |||||  |||  :||   |||||   ||:|:|:|  : |::| |: |||
Db         67 GGFLDDVAGFDAEFFGISEREATSIDPQQRLLLETSWEAIEHAGLDPASLAGSSTAVFTG 126

Qy        133 IGPSEYEAALPQATASAEIDAHGGL-------GTMPSVGAGRISYALGLRGPCVAVDTAY 185
              :   :|                  |||        ||  |:|||:: ||| ||  |  |||
Db        127 LTHEDYLVLTTTA---------GGLASPYVVTGLNNSVASGRIAHTLGLHGPAMTFDTAC 177

Qy        186 SSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADG 245
              || |:|||||||:||  ||   |||||  :::|  |   |   |   |:  ||| :| |:|||
Db        178 SSGLMAVHLACRSLHDGERDLALAGGCAVLLEPHACVAASAQGMLSSTGRCHSFDADADG 237

Qy        246 FGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALA 305
              | | ||||:|:||||  |  ||:|| ||:|||:| |||  :  ||:|: :|  | : |||
Db        238 FVRSEGCAMVLLKRLPDALRDGNRIFAVVRGTATNQDGRTETLTMPSEDAQVAVYRAALA 297

Qy        306 DAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYAS 365
                 || :|| |||||||| :|||| |  |||   |||     ||  :|| |:|:|||   ::
Db        298 AAGVQPETVGVVEAHGTGTPIGDPIEYRSLARVYGAG----TPCALGSAKSNMGHSTASA 353

Qy        366 GITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWN--TPRRAGVS 423
                ||:|  :|||:|| :|  ||  |   |   :|  :   |  |  |:||||:  |  ||:|  ||
Db        354 GTVGLIKAILSLRHGVVPPLLHFNRLPDELSDVETGLFVPQAVTPWPNGNDHTPKRVAVS 413

Qy        424 SFGMSGTNAHVVLEEAPAATCTPPAPERPAE----LLVLSARTASALDAQAARLRDHLET 479
              |||||||  |   ::|||||     |  :     ||      |:||: :: ||    |  :|    :|
Db        414 SFGMSGTNVHAIVEEAPAEASAPESSPGDAEVGPRLFMLSSTSSDALRQTARQLATWVEE 473

Qy        480 YPSQCL--GDVAFSLATTRSAMEHRLAVAATSR----EGLRAALDAAAQGQTSPGAVRSI 533
              :    |:    |:|::||  |         | ||| |  :        ||||   |   |     :  |
Db        474 H-QDCVAASDLAYTLARGRRHRPVRTAVVAANLPELVEGLREVADGDALYDAAVG----- 527

Qy        534 ADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEP 593
                   || : ::|:|||:|  ||  |     |     |     :                | |: ||
Db        528 -HGDRGPV-WVFSGQGSQWAAMGTQLL----ASEPVFAATI--------AKLEPVIAAES 573

Qy        594 A-SVDAALLDQTAFT-----QPALFTFEYALAA-LWRSWGVEPELVAGHSIGELVAACVA 646
                 ||  |:   |    |        |||:|    :  ||||   :  :::||   |    |  |||:||   || ||
Db        574 GFSVTEAITAQQTVTGIDKVQPAVFAVQVALAATMEQTYGVRPGAVVGHSMGESAAAVVA 633

Qy        647 GVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADV-AAAVAPHAASVSIAAVNAPDQ 705
               |  ||||| ::  | :|| :   ||| |:| |   | :  :|    | :: | :|
Db        634 GALSLEDAARVICRRSKLMTRIAGAGAMGSVELPAKQVNSELMARGIDDVVVSVVASPQS 693

Qy        706 VVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIV 765
               || |     |   :   |       |     :  :  |  |||  :  |:|:           ::      | :
Db        694 TVIGGTSDTVRDLIARWEQRDVMAREVAVDVASHSPQVDPILDDLAAALADIAPMTPKVP 753

Qy        766 LVSNLSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPA 825
                |      |   ||| :  |   |:|| |:|  |      |  :  |  |   |
Db        754 YYSATLFDPREQPVCDGAYWVDNLRNTVQFAAAVQAAMEDGYRVFADVSPHPLLTHAVEQ 813

Qy        826 C--MPDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQ 883
                     |    | ||  |   :  |   :   |   |    |   : :: |:|: ||  |   |    |
Db        814 TGRSLDMSVAALAGMRREQPLPHGLRGLLTELHRAGAALDYSALYPA-GRLVDAPLPAWT 872

Qy        884 RERYWID------------------------------------------------------ 890
                | :||
Db        873 HARLFIDDDGQEQRAQGACTITVHPLLGSHVRLTEEPERHVWQGDVGTSVLSWLSDHQVH 932

Qy        891 ------------------------------------------------------------ 890

Db        933 NVAALPGAAYCEMALAAAAEVFGEAAEVRDITFEQMLLLDEQTPIDAVASIDAPGVVNFT 992

Qy        891 ------------TKADDAARGDRRAPG--------------------------------- 905
                          | |   || · |     .||
Db        993 VETNRDGETTRHATAALRAAEDDCPPPGYDITALLQAHPHAVNGTAMRESFAERGVTLGA 1052

Qy        906 --------------------------------------------------------AGH-- 908
                                                                       ||
Db       1053 AFGGLTTAHTPRREPRRCWPRSRCPRRSGSSRAPTESTRRCWTLVSSRSARASSRYGHWW 1112

Qy        909 ----------------------------------------DEVEEGG----AVRG---- 919
                                                     | ::|  |       ||||
Db       1113 PAVAVGCAQPACLRAYPQCPLLLHAVDQGLQRRDPRWEADLDVLDEHGTVLLAVRGLRMG 1172

Qy        920 -----------------------------GDRR---------SARLDHP---------- 930
                                          ||          |  :|  |
Db       1173 TGTSERDERDRLVSERLLTLGWQQRALPEVGDGEAGSWLLIDTSNAVDTPDMLASTLTDA 1232

Qy        931 ------------------------------------------------------------ 930

Db       1233 LKSHGPQGTECASLSWSVQDTPPNDQAGLEKLGSQLRGRDGVVIVYGPRVGDPDEHSLLA 1292

Qy        931 -----------------------------------PPESGRREKVEAAGDR--------- 946
                                                  |  :||  |    :|  ||  |
Db       1293 GREQVRHLVRITRELAEFEGELPRLFVVTRQAQIVKPHDSGERANLEQAGLRGLLRVISS 1352

Qy        947 ------------------------------------------------------------ 946

Db       1353 EHPMLRTTLIDVDEHTDVERVAQQLLSGSEEDETAWRNGDWYVARLTPSPLGHEERRTAV 1412

Qy        947 ------PFRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVP-- 998
                    |:::  |||  |        ::|  |:::|:|  ||    :  ::|  ||  |:| :|  |
Db       1413 LDPDHDGMRVQVRRPGDLQTLEFVASDRVPPGPGQIEVAVSMSSINFADVLIAFGRFPII 1472

Qy        999 DDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALS-AGAFATHVTTSAALVLPR 1057
              ||     :|      ||  :      |  :  |||||  |   ||     |    |    |     |   :|  | |  :
Db       1473 DD--REPQ----LGMDFVGVVTAVGEGVTGHQVGDRVGGFSEGGCWRTFLTCDANLAVTL 1526

Qy       1058 PQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVH 1117
                |  |:     :|        |:  ||||  |:  :|::::  |:::|||||:|||||  ||:   |:   |||:
Db       1527 PPGLTDEQAITAATAHATAWYGLNDLAQIKAGDKVLIHSATGGVGQAAISIARAKGAEIF 1586

Qy       1118 ATAGTPEKRAYLESLGVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLL 1177
              ||||  |   |||  |·  :||  :|  ||||    |     :|   |  |  |||:||||:|                ||
Db       1587 ATAGNPAKRAMLRDMGVEHVYDSRSVEFAEQIRRDTDGYGVDIVLNSLTGAAQRAGLELL 1646
```

```
Qy       1178 RSHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAA 1237
                  |||||:|| | | : :||| || | |:|  :||  | : :| ||| ||   :  | |
Db       1647 PFGGRFVEIGKADVYGNTRLGLFPFRRGLTFYYLDLALMSVTQPDRVRELLATVFKLTAD 1706

Qy       1238 GVFTPPPIATLPIARVADAFRSMAQAQHLGKLVL 1271
               || | |      |:|  ||| |:|: |:| |||||
Db       1707 GVLTAPQCTHYPLAEAADAIRAMSNAEHTGKLVL 1740


RESULT 6
PPSB_MYCTU
ID   PPSB_MYCTU     STANDARD;      PRT;  1538 AA.
AC   Q10978; O53234;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsB.
GN   PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- SIMILARITY: Contains 1 acyl carrier domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL021070; CAA15929.1; -.
DR   EMBL; AE007122; AAK47329.1; -.
DR   PIR; E70874; E70874.
DR   TIGR; MT3002; -.
DR   TubercuList; Rv2932; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; NADP; Phosphopantetheine; Complete proteome.
FT   DOMAIN     1425   1495       ACYL CARRIER (ACP).
FT   BINDING    1458   1458       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1538 AA;  162527 MW;  B55E2A2042AD00CC CRC64;

  Query Match             22.1%;  Score 1592.5;  DB 1;  Length 1538;
  Best Local Similarity   31.2%;  Pred. No. 2.2e-76;
  Matches  488;  Conservative  207;  Mismatches  600;  Indels  271;  Gaps   44;

Qy          3 DRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDP 62
              ||    |  :|:|:|| || || |        || |    |: :  |||:|||| |:: |||
Db         25 DRVSRIAVAEPVAVVGIGCRFPGDVDGPESFWDFLVAGRNAISTVPADRWDAEAFYHPDP 84

Qy         63 DAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSAL 122
                ||:       |: ||| |||  ||||:||||  |||   |:||||  |||||:| |  | :|
Db         85 LTPGRMTTKWGGFVPDVAGFDAEFFGITPREAAAMDPQQRMLLEVAWEALEHAGIPPDSL 144

Qy        123 VGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVD 182
               || |  | :|:   :||:: |   | :    :||: | |    |:  ||||| |||||| ||||
Db        145 GGTRTAVMMGVYFNEYQSML--AASPQNVDAYSGTGNAHSITVGRISYLLGLRGPAVAVD 202

Qy        183 TAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAE 242
              || ||||||||||||||   |     |||||||: | | |  : :|      |:  ||| || |
Db        203 TACSSSLVAVHLACQSLRLRETDLALAGGVSITLRPETQIAISAWGLLSPQGRCAAFDAA 262

Qy        243 ADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKR 302
              |||| ||||  |||||||: |     |||::|||:||||:| || | |:|:| || ::|  |:
Db        263 ADGFVRGEGAGVVVLKRLTDAVRDGDQVLAVVRGSAVNQDGRSNGVTAPNTAAQCDVIAD 322

Qy        303 ALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPE 362
              ||      |     |  || |||||||||  ||||||| :|| | || | |         :|:||||:|| |
Db        323 ALRSGDVAPDSVNYVEAHGTGTVLGDPIEFEALAATYGHGGDACA---LGAVKTNIGHLE 379

Qy        363 YASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGV 422
               |:|| | :|  |::|    || :||     |||    |       |     :|||     |||| |
Db        380 AAAGIAGFIKATLAVQRATIPPNLHFSQWNPAIDAASTRFFVPTQNSPWPTAEGPRRAAV 439

Qy        423 SSFGMSGTNAHVVLEE-APAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYP 481
              ||||: |||||::|: :  |    :     :   ||:: :||    : | |    | | :|
Db        440 SSFGLGGTNAHVIIEQGSELAPVSEGGEDTGVSTLVVTGKTAQRMAATAQVLADWMEGPG 499

Qy        482 SQ-CLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
               .::  : |||  ::   |:           | |  |   | | | | || :|| |  |  |  | |
Db        500 AEVAVADVAHTVNHHRARQATFGTVVARDRAQAIAGLRALAAGQHAPGVV-SHQDGSPGP 558

Qy        541 -LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAA 599
                |:::|:|:|  ||||  |     ||        ||  :|  ::   ||:|:    :
Db        559 GTVFVYSGRGSQWAGMGRQLLADEPAFAAAVAELEPVFVEQAGFSLRDVI---ATGKELV 615

Qy        600 LLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVA 659
               ::|    |  |    :   ||||:|||:|:|:||  |||:||: || |||   :  : : : |
Db        616 GIEQ---IQLGLIGMQLTLTELWRSYGVQPDLVIGHSMGEVAAAVVAGALTPAEGLRVTA 672

Qy        660 ARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIA 719
               | |||  |   || |  :   |   | :|  :   |::   |:| | ||||   : :  :
Db        673 TRARLMAPLSGQGGMALLGLDAAATEALIADY-PQVTVGIYNSPRQTVIAGPTEQIDELI 731

Qy        720 AAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEV 779
              |.: |:    :::  |.|:| |  :  |        :: | |:| ::|            |  :
Db        732 ARVRAQNRFASRVNIEVAPHNPAMDALQPAMRSELADLTPRTPTIGIISTTYADLHTQPI 791

Qy        780 SSPGYWVRHAREVVRFADGVKALHAAGAG------TFVEVGPKSTLLGLVPACMPDA-RP 832
                :|  :  |  |||.   |: :||:|     ||:|:      |     :  || ||
Db        792 FDAEHWATNMRNPVRFQ---QAIASAGSGADGAYHTFIEISAHPLLTQAIADTLEDAHRP 848

Qy        833 A-----LLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVP----------- 876
```

```
                       |:        ||   ||                      |: :     |
Db        849 TKSAAKYLSIGTLQRDADDTV-------------TFRTNLYTADIAHPPHTCHPPEPHPT 895

Qy        877 LPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEGGAVRGGDRRSARLDHP---PPE 933
               :|| |||    :|| |                                 ||     ||
Db        896 IPTTPWQHTHHWIAT-----------------------------------THPSTAAPE 919

Qy        934 SGRREKVEAAGDR----------------PFR--LEIDEPGV----------LDHLVLRV 965
                    ||  |                       | |  :  | |            | | :  |
Db        920 DPGSNKVVVNGQSTSESRALEDWCHQLAWPIRPAVSADPPSTAAWLVVADNELCHELARA 979

Qy        966 TERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEG 1025
                 : |   |    :|   :    :   |    ||   | :  |    | |   ||  | |  ::
Db        980 ADSRVDSLSPPALAAGSDPAALLD---ALRGVDNVLYAPPVPGELLDIESAYQVFHATRR 1036

Qy       1026 VNGLVVGQPVIALSAGAFATHVTTSA---------------------ALVLPRPQALSAI 1064
                :   :|      |:|        :| :|                      :| |  |:    |
Db       1037 LAAAMVASSATAISPPKLFI-MTRNAQPISEGDRANPGHAVLWGLGRSLALEHPEIWGGI 1095

Qy       1065 ---------EAAAMPVAYLTAWYALD------------RIARLQ----PGERVLIHA--- 1096
                          | |    |  ||| :   |                  : |||     ||: |  ::|
Db       1096 IDLDDSMPAELAVRHV--LTAAHGTDGEDQVVYRSGARHVPRLQRRTLPGKPVTLNADAS 1153

Qy       1097 -----ATGGVGLAAVQWAQHVGAE-VHATAGTP----EKRAYLESLGVRYVS-------- 1138
                   ||| :     ::       :||: :  |   |         |  : |      ::
Db       1154 QLVIGATGNIGPHLIRQLARMGAKTIVAMARKPGALDELTQCLAATGTDLIAVAADATDP 1213

Qy       1139 ---DSRSDRFVADVRAWTG-----GEGVDVVLNSLSGELIDKSFN-----LLRSHGRFVE 1185
                  :  |||   ::         |            |      :|: :: :  :    |         |    | |  ::
Db       1214 AAMQTLFDRFGTELPPLEGIYLAAFAGRPALLSEMTDDDVTTMFRPKLDALALLHRRSLK 1273

Qy       1186 ---------------LGKR---------DCYADNQLGLRPFLR---------NLSFSLVDLR 1214
                              ||  |                :   |:   |  |   :            |   ||  |::
Db       1274 SPVRHFVLFSSVSGLLGSRWLAHYTATSAFLDSFAGARRTMGLPATVVDWGLWKSLADVQ 1333

Qy       1215 --GMMLERPARVRALLEEL-LGLIAAGVFTPPPIATLPIAR----VADAFRSMAQAQHLG 1267
                   :    : :: : :|: :|  :    :       :||: :|        :|  |:|:         : :
Db       1334 KDATQISAESGLQPMADEVAIGALPLVMNPDAAVATVVVAADWPLLAAAYRTRGALRIVD 1393

Qy       1268 KLVLT---LGDPEVQIRIPTHAGAGPSTGDRDLL-DRLASAAPAARAAALEAFLRTQVSQ 1323
               |:       :|   |   |    |     |   :   |:     |:|  |  : : |
Db       1394 DLLPAPEDVGKGESEFR--TSLRSCPAEKRRDMLFDHVGALAATVMG------------ 1438

Qy       1324 VLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLD 1383
                   |     :   |  :  :|||||||:|   |:   :    ||      |   :    :     :| |
Db       1439 --MPPTEPLDPSAGFFQLGMDSLMSVTLQRALSESLGEFLPASVVFDYPTV----YSLTD 1492

Qy       1384.ALATAL 1389
                 ||| |
Db       1493 YLATVL 1498


RESULT 7
PPSA_MYCTU
ID   PPSA_MYCTU     STANDARD;      PRT;  1876 AA.
AC   Q10977;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsA.
GN   PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC       similarity).
CC   -!- SIMILARITY: Contains 2 acyl carrier domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z74697; CAA98988.1; -.
DR   EMBL; AE007122; AAK47328.1; -.
DR   PIR; C70749; C70749.
DR   TIGR; MT3000; -.
DR   TubercuList; Rv2931; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
KW   Phosphopantetheine; Complete proteome.
FT   DOMAIN        7     80       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1764   1833       ACYL CARRIER (ACP) 2.
FT   NP_BIND    1498   1503       NADP (POTENTIAL).
FT   BINDING      62     62       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    273    273       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    720    720       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1796   1796       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    624    624       D -> E (IN REF. 2).
FT   CONFLICT    877    877       R -> H (IN REF. 2).
FT   CONFLICT   1323   1323       G -> S (IN REF. 2).
SQ   SEQUENCE   1876 AA;  198834 MW;  D9783DBD48792110 CRC64;

  Query Match             22.0%;  Score 1583;  DB 1;  Length 1876;
  Best Local Similarity   29.2%;  Pred. No. 8.9e-76;
  Matches  532;  Conservative  205;  Mismatches  622;  Indels  460;  Gaps   49;

Qy          2 ADRPIERAA----EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAW 57
               :|   ::| |       ::|||:||  || |||:         |    |     | :: :|| :||
Db         88 SDAAVKRGARNSLDEPIAVVGMGCRFPGGISCPEALWDFLCERRSSISQVPPQRWQPFEG 147

Qy         58 FDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAI 117
                 |:   |                |||  |:  |||  ||  |||  ||  :|||   ||||||  |||||:|  |
Db        148 GPPEVAAALARTTRWGSFLPDIDAFDAEFFEISPSEADKMDPQQRLLLEVAWEALEHAGI 207
```

```
Qy      118 APSALVGTETGVFIGIGPSEYEAALPQATAS-AEIDAHGGLGTMPSVGAGRISYALGLRG 176
Db      208 PPGTLRRSATGVFAGACLSEYGA---MASADLSQVDGWSNSGGAMSIIANRLSYFLDLRG 264

Qy      177 PCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRC 236
Db      265 PSVAVDTACSSSLVAIHLACQSLRTQDCHLAIAAGVNLLLSPAVFRGFDQVGALSPTGQC 324

Qy      237 KAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQ 296
Db      325 RAFDATADGFVRGEGAGVVVLKRLTDAQRDGDRVLAVICGSAVNQDGRSNGLMAPNPAAQ 384

Qy      297 EIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKT 356
Db      385 MAVLRAAYTNAGMQPSEVDYVEAHGTGTLLGDPIEARALGTVLGRGRPEDSPLLIGSVKT 444

Qy      357 NLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNT 416
Db      445 NLGHTEAAAGIAGFIKTVLAVQHGQIPPNQHFETANPHIPFTDLRMKVVDTQTEWPATGH 504

Qy      417 PRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPERPA-ELLVLSARTASALDAQAARLRD 475
Db      505 PRRAGVSSFGFGGTNAHVVIEQGQEVRPAPGQGLSPAVSTLVVAGKTMQRVSATAGMLAD 564

Qy      476 HLETYPSQ-CLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIA 534
Db      565 WMEGPGADVALADVAHTLNHHRSRQPKFGTVVARDRTQAIAGLRALAAGQHAPGVVNP-A 623

Qy      535 DSSRGK-LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEP 593
Db      624 DGSPGPGTVFVYSGRGSQWAGMGRQLLADEPAFAAAVAELEPVFVEQAGFSLHDVL---- 679

Qy      594 ASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLED 653
Db      680 --ANGEELVGIEQIQLGLIGMQLALTELWCSYGVRPDLVIGHSMGEVAAAVVAGALTPAE 737

Qy      654 AVFLVAARGRLMQALPAGG--AMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGA 711
Db      738 GLRVTATRSRLMAPLSGQGGMALLELDAPTTEALIADFP---QVTLGIYNSPRQTVIAGP 794

Qy      712 GQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLS 771
Db      795 TEQIDELIARVRAQNRFASRVNIEVAPHNPAMDALQPAMRSELADLTPRTPTIGIISTTY 854

Qy      772 GKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAG------TFVEVGP----KSTLLG 821
Db      855 ADLHTQPVFDAEHWATNMRNPVRFQ---QAIASAGSGADGAYHTFIEISAHPLLTQAIID 911

Qy      822 LVPACMPDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFP-----SGGRRVP 876
Db      912 TLHSAQPGARYTSLGTLQRDTDDVVTFRTNLNKAHTI---------HPPHTPHPPEPHPP 962

Qy      877 LPTYPWQRERYWIDTKADDAARGDRRAPGAG-----HDEVEEGGAVRGGDRRSARL--DH 929
Db      963 IPTTPWQHTRHWITTKYPAGSVGS--APRAGTLLGQHTTVATVSASPPSHLWQARLAPDA 1020

Qy      930 PPPESGRR----EKVEAA---------------------------GDRPFRLEIDEPGV 957
Db     1021 KPYQGGHRFHQVEVVPASVVLHTILSAATELGYSALSEVRFEQPIFADRPRLIQV----V 1076

Qy      958 LDHLVLRVTERRA---------------------------------------------P 971
Db     1077 ADNRAISLASSPAAGTPSDRWTRHVTAQLSSSPSDSASSLNEHHRANGQPPERAHRDLIP 1136

Qy      972 GLGEVEIAVDAAGLSF-----------NDVQLALGMVPDDLPGKPNPPLLGGECAGRIV 1020
Db     1137 DLAELLAMRGIDGLPFSWTVASWTQHSSNLTVAIDL-PEALPEGSTGPLL---DAAVHLA 1192

Qy     1021 AVGEGVNG-LVVGQPVIALSAGAFATHVTTSAAL-----------VLPRPQALSAIEAAA 1068
Db     1193 ALSDVADSRLYVPASIEQISLGDVVTGPRSSVTLNRTAHDDDGITVDVTVAAHGEVPSLS 1252

Qy     1069 M-PVAYLTAWYALDRIARLQPGERVLIHAATGGVG--------LAAVQW-------AQHV 1112
Db     1253 MRSLRYRALDFGLD-VGRAQP------PASTGPVEAYCDATNFVHTIDWQPQTVPDATHP 1305

Qy     1113 GAE--VH-----------ATAGTPEKRAYLESLGVRYVSDSRSDRFVAD---------- 1148
Db     1306 GAEQVTHPGPVAIIGDDGAALCETLEGAGYQPAVMSDGVSQARYVVYVADSDPAGADETD 1365

Qy     1149 ----VRAWTGGEG--------------------------------------------- 1157
Db     1366 VDFAVRICTEITGLVRTLAERDADKPAALWILTRGVHESVAPSALRQSFLWGLAGVIAAE 1425

Qy     1158 --------VDVVLNSLSGELIDKSFNLLRSHGRFVELGKRD------------------ 1190
Db     1426 HPELWGGLVDLAINDDLGEFGPALAELLAKPSKSI-LVRRDGVVLAPALAPVRGEPARKS 1484

Qy     1191 --CYAD---------NQLGL----------------------RPFLRNLSFSLVD------ 1212
Db     1485 LQCRPDAAYLITGGLGALGLLMADWLADRGAHRLVLTGRTPLPPRRDWQLDTLDTELRRR 1544

Qy     1213 --------LRGMMLERPAR-------VRALL--------EELLGLI-------------- 1235
Db     1545 IDAIRALEMRGVTVEAVAADVGCREDVQALLAARDRDGAAPIRGIIHAAGITNDQLVTSM 1604

Qy     1236 ------------------------------------AAGVFTPPP----------IATLP 1249
Db     1605 TGDAVRQVMWPKIGGSQVLHDAFPPGSVDFFYLTASAAGIFGIPGQGSYAAANSYLDALA 1664

Qy     1250 IARVADAFRSM-------------AQAQHLGKLVLTLGDPE------------------- 1277
Db     1665 RARRQQGCHTMSLDWVAWRGLGLAADAQLVSEELARMGSRDITPSEAFTAWEFVDGYDVA 1724

Qy     1278 --VQIRIPTHAGAGPSTGDRDLLDRLASAAPAAR--AAALEAFLRTQVSQVLRTPEIKVG 1333
Db     1725 QAVVVPMPAPAGADGSGANAYLLPARNWSVMAATEVRSELEQGLRRIIAAELRVPEKELD 1784

Qy     1334 AEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATALSLER 1393
Db     1785 TDRPFAELGLNSLMAMAIRREAEQFVGIELSATMLFNHPTVKSLASYL---------AKR 1835

Qy     1394 VAAENLRAGVQNDFVSSGA 1412
Db     1836 VAPHDVSQDNQISALSSSA 1854
```

RESULT 8
WA_EMENI

```
ID   WA_EMENI       STANDARD;      PRT;  1986 AA.
AC   Q03149;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Conidial green pigment synthase (EC 2.3.1.-).
GN   WA.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=162425;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93101122; PubMed=1465094;
RA   Mayorga M.E., Timberlake W.E.;
RT   "The developmentally regulated Aspergillus nidulans wA gene encodes a
RT   polypeptide homologous to polyketide and fatty acid synthases.";
RL   Mol. Gen. Genet. 235:205-212(1992).
CC   -!- FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN
CC       INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED
CC       BY CONIDIAL LACCASE TO FORM THE GREEN PIGMENT IN MATURE
CC       ASEXUAL SPORES (CONIDIA).
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- PATHWAY: Conidial green pigment biosynthesis.
CC   -!- SIMILARITY: Contains 2 acyl carrier domains.
```

```
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X65866; CAA46695.1; -.
DR   PIR; S28353; S28353.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN      529    582       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   DOMAIN      991   1024       ACYL/MALONYL TRANSFERASES
FT                                (BY SIMILARITY).
FT   DOMAIN     1650   1719       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1772   1841       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    548    548       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE   1001   1001       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1682   1682       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1804   1804       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1986 AA;  216634 MW;  74EF0940FF40EE9A CRC64;

  Query Match             16.4%;  Score 1180;  DB 1;  Length 1986;
  Best Local Similarity   26.9%;  Pred. No. 1.6e-54;
  Matches  419;  Conservative  242;  Mismatches  635;  Indels  264;  Gaps   53;

Qy          8 RAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGK 67
Db        373 RAELSKIAIIGMSGRFP-EADSPQDFWNLLYKGLDVHRKVPEDRWDADAHVDLTGTATNT 431

Qy         68 TPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTE- 126
Db        432 SKVPYGCWIREPGLFDPRFFNMSPREALQADPAQRLALLTAYEALEGAGFVPDSTPSTQR 491

Qy        127 --TGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184
Db        492 DRVGIFYGMTSDDYR----EVNSGQDIDTYFIPGGNRAFTPGRINYYFKFSGPSVSVDTA 547

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244
Db        548 CSSSLAAIHLACNSIWRNDCDTAITGGVNILTNPDNHAGLDRGHFLSRTGNCNTFDDGAD 607

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304
Db        608 GYCRADGVGTVVLKRLEDALADNDPILGVINGAYTNHSAEAVSITRPHVGAQAFIFKKLL 667

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLG--RDVATPLLIGSVKTNLGHPE 362
Db        668 NEANVDPKNISYIEMHGTGTQAGDAVEMQSVLDVFAPDHRRGPGQSLHLGSAKSNIGHGE 727

Qy        363 YASGITGLLKVVLSLQHGQIPAHLHAQA-LNPRISWGDLRLTVTRARTPWPDWNTP---- 417
Db        728 SASGVTSLVKVLLMMKENMIPPHCGIKTKINHNFPTDLAQRNVHIALQP-TAWNRPSFGK 786

Qy        418 RRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE----RPAELLVLSARTASAL----DAQ 469
Db        787 RQIFLNNFSAAGGNTALLLEDGPVSD-----PEGEDKRRTHVITLSARSQTALQNNIDAL 841

Qy        470 AARLRDHLETY---PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQ-- 524
Db        842 CQYISEQEKTFGVKDSNALPSLAYTTTARRIHHPFRVTAIGSSFQEMRDSLIASSRKEFV 901

Qy        525 ----TSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDL--CVRLFN 578
Db        902 AVPAKTPG------------IGFLFTGQGAQYAAMGKQLYEDCSHFRSAIEHLDCISQ-G 948

Qy        579 QELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIG 638
Db        949 QDLPSILPLVDGSLPLSELSPVVVQLGTT-----CVQMALSSFWASLGITPSFVLGHSLG 1003

Qy        639 ELVAACVAGVFSLEDAVFLVAARGRLM--QALPAGGAMVSIEAPEADVAAAVAPHAASVS 696
Db       1004 DFAAMNAAGVLSTSDTIYACGRRAQLLTERCQPGTHAMLAIKAPLVEVKQLL--NEKVHD 1061

Qy        697 IAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAES 756
Db       1062 MACINSPSETVISGPKSSIDELSRACSEKGLKSTILTVPYAFHSAQVEPILEDLEKALQG 1121

Qy        757 VSYRRPSIVLVSNLSGKACTD---EVSSPGYWVRHAREVVRFADGVKALHAAGAG----T 809
Db       1122 ITFNKPSVPFVSALLGEVITEAGSNILNAEYLVRHCRETVNFLSAFEAVRNAKLGGDQTL 1181

Qy        810 FVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGL-- 867
Db       1182 WLEVGPHTVCSGMVKATL-GPQTTTMASLRRDEDTWKVLSNSLSSLYLAGVDINWKQYHQ 1240

Qy        868 -FPSGGRRVPLPTYPWQRERYWID-------TKADDAARGDRRAPGAGHDEVEEGGAVRG 919
Db       1241 DFSSSHRVLPLPTYKWDLKNYWIPYRNNFCLTKGSSMS----AASASLQPTFLTTSAQRV 1296

Qy        920 GDRRSARL--------DHPPPESGR---REKVEAAGDRPFRLEIDEPGVLDHLVLRVTER 968
Db       1297 VESRDDGLTATVVVHNDIADPDLNRVIQGHKVNGAALCPSSLYADSAQTLAEYLI---EK 1353

Qy        969 RAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNG 1028
Db       1354 YKPEL-------KGSGLDVCNVTV-----------PKPLIAKTGKEQFRISATANWVDK 1394

Qy       1029 LVVGQPVIALSAG---------------AFATHVTTSAALVLPRPQALSAIEAAAMP--- 1070
Db       1395 HVSVQVFSVTAEGKKLIDHAHCEVKLFDCMAADLEWKRGSYLVK-RSIELLENSAVKGDA 1453

Qy       1071 ------VAYLTAWYALDRIARLQPGERVLI----HAATGGVGLAAVQ--------WAQ-- 1110
Db       1454 HRLRRGMVYKLFSALVDYDENYQSIREVILDSEHHEATALVKFQAPQANFHRNPYWIDSF 1513

Qy       1111 -HV-GAEVHATAGTPEKRAYLESLGVRYVSDSRS--DRFVADV-------------RAWT 1153
Db       1514 GHLSGFIMNASDGTDSK----SQVFVNHGWDSMRCLKKFSADVTYRTYVRMQPWRDSIWA 1569

Qy       1154 GG----EGVDVV-------LNSLSGELIDKSF----------NLLRSHG--RFVELGKRD 1190
Db       1570 GNVYIFEGDDIIAVFGGVKFQALSRKILDIALPPAGLSKAQTSPIQSSAPQKPIETAKPT 1629

Qy       1191 CYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAG-----VFTPPPI 1245
Db       1630 SRPAPPVTMKSFVKKSAGPSVVVRAL---------NILASEVGLSESDMSDDLVFADYGV 1680

Qy       1246 ATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPT-------------HAGAGPST 1292
Db       1681 DSLLSLTVTGKYRE--------ELNLDM-DSSVFIEHPTVGDFKRFVTQLSPSVASDSSS 1731

Qy       1293 GDR--------DLLDRLAS-AAPAARAAALEAFLRTQVSQVLR------TPEIKVGAEAL 1337
Db       1732 TDRESEYSFNGDSCSGLSSPASPGTVSPPNEKVIQIHENGTMKEIRAIIADEIGVSADEI 1791

Qy       1338 -----FTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATALSLE 1392
Db       1792 KSDENLNELGMDSLLSLTVLGKIRESLDMDL--------PGEFFIENQTLDQIETALDLK 1843
```

```
RESULT 9
MSAS_PENPA
ID   MSAS_PENPA     STANDARD;      PRT;  1774 AA.
AC   P22367;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   6-methylsalicylic acid synthase (EC 2.3.1.165) (6-MSAS).
OS   Penicillium patulum (Penicillium griseofulvum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5078;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=DSM 62862;
RX   MEDLINE=91006137; PubMed=2209605;
RA   Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E.;
RT   "The multifunctional 6-methylsalicylic acid synthase gene of
RT   Penicillium patulum. Its gene structure relative to that of other
RT   polyketide synthases.";
RL   Eur. J. Biochem. 192:487-498(1990).
CC   -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC       It catalyzes a total of 11 steps by seven different component
CC       enzymes, in the biosynthesis of the antibiotic patulin.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH = 6-
CC       methylsalicylate + 4 CoA + 3 CO(2) + NADP(+).
CC   -!- PATHWAY: Patulin biosynthesis.
CC   -!- SUBUNIT: HOMOMULTIMER.
CC   -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC   -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC       FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC       THIOLASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X55776; CAA39295.1; -.
DR   PIR; S13178; S13178.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Multifunctional enzyme; Transferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine.
FT   DOMAIN      186    238       ACYLTRANSFERASE (AT).
FT   DOMAIN      642    676       ACETYL/MALONYL TRANSFERASES.
FT   DOMAIN     1403   1450       2-OXOACYL REDUCTASE.
FT   DOMAIN     1700   1769       ACYL CARRIER (ACP).
FT   NP_BIND    1419   1424       NADP (POTENTIAL).
FT   ACT_SITE    204    204       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    653    653       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1732   1732       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1774 AA;  190732 MW;  05ED5DD10863F938 CRC64;

  Query Match            16.3%;  Score 1177;  DB 1;  Length 1774;
  Best Local Similarity  25.6%;  Pred. No. 2e-54;
  Matches  458;  Conservative  210;  Mismatches  655;  Indels  466;  Gaps   56;

Qy         14 IAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKTPVTRA 73
Db         35 VAVVGMACRVAGGNHNPELLWQSLLSQKSAMGEIPPMRWEPYYRRDARNEKFLKNTTSRG 94

Qy         74 SFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGVFIGI 133
Db         95 YFLDRLEDFDCQFFGISPKEAEQMDPQQRVSLEVASEALEDAGIPAKSLSGSDTAVFWGV 154

Qy        134 GPSEYEAA----LPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSSL 189
Db        155 NSDDYSKLVLEDLP------NVEAWMGIGTAYCGVPNRISYHLNLMGPSTAVDAACASSL 208

Qy        190 VAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGRG 249
Db        209 VAIHHGVQAIRLGESKVAIVGGVNALCGPGLTRVLDKAGAISSDGSCKSFDDDAHGYARG 268

Qy        250 EGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAGC 309
Db        269 EGAGALVLKSLHRALLDHDNVLAVIKGSAVCQDGKTNGIMAPNSVAQQLAANNALSAANI 328

Qy        310 AASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITG 369
Db        329 DPHTVRYVEAHATSTPLGDPTEISAIASVYGADRPADDPCYIGSIKPNIGHLEAGAGVMG 388

Qy        370 LLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSG 429
Db        389 FIKAVLAIQKGVLPPQANLTKLNSRIDWKTAGVKVVQEATPWPESDPIRRAGVCSYGYGG 448

Qy        430 TNAHVVLEE-APAATCTP---PAPERPAELLVLSARTASALDAQAARLRDHLETY-PSQC 484
Db        449 TVSHAVIEEFSPILQPDPLGNGAVSGPG-LLLLSGPQEKRLALQAKTLRDWMTAEGKDHN 507

Qy        485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQG--QTSPGAVRSIADSSRGKLA 542
Db        508 LSDILTTLATRRDHHDYRAALVVDDYRDAEQVLQSLANGVDHTFTTQSRVLGSDISKDVV 567

Qy        543 FLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLD 602
Db        568 WVFSGHGAQWPDMGKQL-----IHNPVFFAAI--------QPLDELIQAEIGLSPIELLR 614

Qy        603 QTAF-----TQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFL 657
Db        615 TGDFESSDRVQILTYVMQIGLSALLQSNGITPQAVIGHSVGEIAASVVAGALSPAEGALI 674

Qy        658 VAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHA 717
Db        675 VTRRALLYRQVMGKGGMILVNLPSAETEEILGSRSDLV-VAIDSSPSSCVVAGDKELVAE 733

Qy        718 IAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRV-AESVSYRRPSIVLVSNLSGKACT 776
Db        734 TAEALKARGVKTFTVKSDIAFHSPTLNGLVDPLRDVLAETLSPVSPNVKLYSTALADPRG 793

Qy        777 DEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDAR----- 831
Db        794 QDLRDVEYWAGNMVNRVRLTSAVKAAVEDGYRLFLEVSTHPVVSHSINETLMDAGMEDFA 853

Qy        832 --PALLASSRAGRDEPAT--VLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERY 887
Db        854 VIPTLL------RKKPTEKHILHSIAQLHCRGAEVNWAAQMP-GRWATGVPTTTWMHKPI 906

Qy        888 WIDTKADDAARGDRRAP---GAGHDEVEEGG------AVRGGDR--RSARLDH---PPPE 933
Db        907 W---------RKIETAPLHTGLTHD-VEKHTLLGQRIPVPGTDTYVYTTRLDNDTKPFPG 956

Qy        934 S---GRREKVEAAG-------------------------DRPFRLEI------------ 952
Db        957 SHPLHGTEIVPAAGLINTFLKGTGGQMLQNVVLRVPVAINAPRSVQVVVQQDQVKVVSRL 1016

Qy        953 --DEPGVLDHLVLRVT------ERRAPGLGE-----------------------VEIAV 980
Db       1017 IPSEPSQLDDDASWVTHTTAYWDRKVAGSEDRIDFAAVKSRLVTKLADNFSIDYLDKVGV 1076

Qy        981 DAAGLSF-------NDVQLALGM-----VPDDLP-------------------GK---PNP 1007
Db       1077 SAMGFPWAVTEHYRNDKEMLARVDVNPAISGDAPLPWDSSSWAPVLDAATSVGSTIFPTP 1136
```

```
Qy       1008 PLLL----------------------------------------GGECAGRIVAVG---- 1023
              | :                                         ||   :   |:
Db       1137 ALRMPAQIERVEVFTSQDPPKISWLYVQEASDSVPTSHVSVVSEAGEVLAKFTAMRFSEI 1196

Qy       1024 EGVNGL------VVGQPVIALSAGAFATHVTTSAALVLPRPQALS-AIEAAAMP------ 1070
              ||  |:     :|  |   ||        |     :   ::|  | |  : |: ||::|
Db       1197 EGTPGVSGSMESLVHQ--IAWPPATPAEEPLSIETVILVSPDATTRALYAASLPTRVNSF 1254

Qy       1071 ----------------------VAYLT---------------AWYALDRIARLQPGE-- 1090
                                    | |:               |   | :  |    |
Db       1255 QFSSTQEFFSNASSLPLEKGTVVTYIPGEVASLAEVPAASESFTWNLLELIKFTVNGSLP 1314

Qy       1091 -RVLIHAATGGVGLAAVQWAQ-------HVGAEVHATAGT----------PEKRAYLESL 1132
               :|    |  | |     ||         | |  |  ||                 |::
Db       1315 IKVFTLTANIGEGQTPTALAQSPLYGLARVIASEHPDLGTLIDVEEPVIPLSTMRYIQGA 1374

Qy       1133 GVRYVSD--SRSDRFVADVR--------------------AWTGGEGV------------ 1158
               :   ::|  :|: || :  |                    ||| ||
Db       1375 DIIRINDGIARTSRFRSLPRNKLLPASEGPRLLPRPEGTYLITGGLGVLGLEVADFLVEK 1434

Qy       1159 -------------------DVVLNSLSGELIDKSFNLLRSHGRFV-----ELGKRDCYAD 1194
                                 |  |     :        || |  |          :: |  |
Db       1435 GARRLLLISRRALPPRRTWDQVSEDLQPTI--AKIRLLESRGASVHVLPLDITKPDAVEQ 1492

Qy       1195 NQLGLRPFLRNLSF----SLVDLRG-------MMLERPARVRALLEELLGLIAAGVFTPP 1243
                   |   |   ||          :|    |    |  |  |  |   :: :|  :|      ||
Db       1493 ----LTTALDRLSLPSVQGVVHAAGVLDNELVMQTTRDAFNRVLAPKIAGALALHEVFPP 1548

Qy       1244 P-----------------------------IATLPI--ARVADAF--------------- 1257
                                             : ||    ||: ||
Db       1549 KSVDFFVMFSSCGNLVGFTGQASYGSGNAFLDTLATHRARLGDAAVSFQWTSWRGLGMGA 1608

Qy       1258 --------------------RSMAQAQHLGKL--------------------VLTLGDPE 1277
                                  :  |  |||  |                        |   |   |
Db       1609 STDFINAELESKGITDVTRDEAFAAWQHLAKYDMDHGVVLRSRAFEDGEPIPVSILNDIA 1668

Qy       1278 VQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEAL 1337
               |:  |:  |   :      |:           |  : ::  |   :  |  |:  :|  |::||:      |  ::|
Db       1669 VR-RVGTVSNTSPAAAGSS--DAVPTSGPELK-AYLDEKIRGCVAKVLQMTAEDVDSKAA 1724

Qy       1338 FTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALA 1386
                 ||:||:|  |  ||  :::  :||:  :    |    :  |  ::  ||    :  ||
Db       1725 LADLGVDSVMTVTLRRQLQLTLKIAVPPTLTWSHPTVSHLAVWFAEKLA 1773
```

```
RESULT 10
PKS1_ASPPA
ID   PKS1_ASPPA     STANDARD;      PRT;  2109 AA.
AC   Q12053;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Aflatoxin biosynthesis polyketide synthase (PKS).
GN   PKSL1.
OS   Aspergillus parasiticus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5067;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2999;
RX   MEDLINE=96042102; PubMed=7592391;
RA   Feng G.H., Leonard T.J.;
RT   "Characterization of the polyketide synthase gene (pksL1) required
RT   for aflatoxin biosynthesis in Aspergillus parasiticus.";
RL   J. Bacteriol. 177:6246-6254(1995).
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC       AFLATOXIN FROM HEXANOYL COA AND SEVEN MALONATES.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- PATHWAY: Aflatoxin biosynthesis; first step.
CC   -!- SIMILARITY: Contains 1 acyl carrier domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L42766; AAC41675.1; -.
DR   EMBL; L42765; AAC41674.1; -.
DR   PIR; T17490; T17490.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
KW   Transferase; Acyltransferase; Phosphopantetheine;
KW   Multifunctional enzyme.
FT   DOMAIN      374    805       BETA-KETOACYL SYNTHASE.
FT   DOMAIN        ?      ?       ACYL/MALONYL TRANSFERASES.
FT   DOMAIN     1714   1785       ACYL CARRIER (ACP).
FT   DOMAIN        ?   2109       THIOESTERASE.
FT   ACT_SITE    543    543       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    993    993       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1746   1746       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1937   1937       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2109 AA;  230715 MW;  CB701372A16D8551 CRC64;

  Query Match             16.3%;  Score 1173.5;  DB 1;  Length 2109;
  Best Local Similarity   27.2%;  Pred. No. 3.8e-54;
  Matches  409;  Conservative  222;  Mismatches  625;  Indels  247;  Gaps   54;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKT 68
              | :  :|||  |  |  |         | || ||     |       ||  |||       ||   |   |
Db        369 AGKCKLAIVSMSGRFPESPTTES-FWDLLYKGLDVCKEVPRRRWDINTHVDPSGKARNKG 427

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTE-- 126
                      :|       ||  ||||||:|| :||||  |: |     :||:|  |  :  |        |:
Db        428 ATKWGCWLDFSGDFDPRFFGISPKEAPQMDPAQRMALMSTYEAMERAGLVPDTTPSTQRD 487

Qy        127 -TGVFIGIGPSEYEAALPQATASAEIDAH----GGLGTMPSVGAGRISYALGLRGPCVAV 181
                |||  |:  :::       :    :   ||  :       |   |  :|      |||::       ||
Db        488 RIGVFHGVTSNDW----METNTAQNIDTYFITGGNRGFIP----GRINFCFEFAGPSYTN 539

Qy        182 DTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSA 241
               ||| ||||  |:||||  ||    |:|  ||:|||  :::  :|          |  |      |:|  | |  ||  :
Db        540 DTACSSSLAAIHLACNSLWRGDCDTAVAGGTNMIYTPDGHTGLDKGFFLSRTGNCKPYDD 599

Qy        242 EADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLK 301
              :|||: |  ||    | :|||  | ||  | ||  | ||  | ||   :   ||    |  :|  |:  :|    :
Db        600 KADGYCRAEGVGTVFIKRLEDALADNDPILGVILDAKTNHSAMSESMTRPHVGAQIDNMT 659

Qy        302 RALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYG---LGRDVATPLLIGSVKTNL 358
                ||    |    :      |:|  |||||  :||  :|::::  :|:           |      || :|| |  |:
Db        660 AALNTTGLHPNDFSYIEMHGTGTQVGDAVEMESVLSVFAPSETARKADQPLFVGSAKANV 719

Qy        359 GHPEYASGITGLLKVVLSLQHGQIPAHL---HAQALNPRI-SWGDLRLTVTRARTPWPDW 414
               || |    ||:|  |:||::  :||   || |              :|           |   :  :        |||
Db        720 GHGEGVSGVTSLIKVLMMMQHDTIPPHCGIKPGSKINRNFPDLGARNVHIAFEPKPWPRT 779

Qy        415 NTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLR 474
               :||||     :::|    :|  |     :::|:||                |  :  ::  |||       :::         ||
Db        780 HTPRRVLINNFSAAGGNTALIVEDAPERHWPTEKDPRSSHIVALSAHVGASMKTNLERLH 839

Qy        475 DHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQ---GQTSPGAVR 531
```

```
                 :|    |    |  ::::     |     ||::|    | |  :     |: |  |    | : |
Db     840 QYLLKNPHTDLAQLSYTTTARRWHYLHRVSVTGASVEEVTRKLEMAIQNGDGVSRP---- 895

Qy     532 SIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRL--------FNQELDR 583
                   |: |: | |||||:|   ||: :|| : :|||  :    ||         |
Db     896 ----KSKPKILFAFTGQGSQYATMGKQVYDAYPSFREDLEKFDRLAQSHGFPSFLHVCTS 951

Qy     584 PLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAA 643
            |  :|    |     |        | |:   : ||   |  |:|: |:: |||:|| |
Db     952 PKGDVEEMAPVVV-----------QLAITCLQMALTNLMTSFGIRPDVTVGHSLGEFAAL 1000

Qy     644 CVAGVFSLEDAVFLVAARGRLMQALPAGG--AMVSIEAPEADVAAAVAPHAASVSIAAVN 701
              ||| |   | |:|| |  |:|      | ||::::|     ::  :  |      :| :|
Db    1001 YAAGVLSASDVVYLVGQRAELLQERCQRGTHAMLAVKATPEALSQWIQDH--DCEVACIN 1058

Qy     702 APDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRR 761
             |:  |::|  : |  :  ||   |  :    |  :  ||||  : |:|: |  :|:  :: :
Db    1059 GPEDTVLSGTTKNVAEVQRAMTDNGIKCTLLKLPFAFHSAQVQPILDDFEALAQGATFAK 1118

Qy     762 PSIVLVSNLSGKACTDE-VSSPGYWVRHAREVVRFADGVKALHAAG----AGTFVEVGPK 816
            | ::::| |       :: | :|  |  :| |   |   |  :::     |          :|:|||
Db    1119 PQLLILSPLLRTEIHEQGVVTPSYVAQHCRHTVDMAQALRSAREKGLIDDKTLVIELGPK 1178

Qy     817 STLLGLVPACMPDARPAL--LASSRAGRDEPATVLEALGGLWAVGGLVSWA---GLFPSG 871
             : |:|   :   |      || ::|      :: :  |:    |   :: :|        | |
Db    1179 PLISGMVKMTLGDKISTLPTLAPNKA---IWPSLQKILTSVYTGGWDINWKKYHAPFASS 1235

Qy     872 GRRVPLPTYPWQRERYWIDTKAD---DAARGDRRAPGAGHDEVEEGGAVRGGDRRSARLD 928
             : |  ||:|  |   :  |:|  |          : |  :      ||:         ::   |  :
Db    1236 QKVVDLPSYGWDLKDYYIPYQGDWCLHRHQQDCKCAAPGHE-------IKTADYQ----- 1283

Qy     929 HPPPESGRREKVEAAGDRPFRLEIDE---PGVLDHLVL-RVTERRAPGLGEVEIAVDAAG 984
             ||||          || |:|:  :     | :       | |||  |       ||      |
Db    1284 -VPPES--------TPHRPSKLDPSKEAFPEIKTTTTLHRVVEETTKPLGAT--LVVETD 1332

Qy     985 LSFNDVQ-LALGMVPDDLPGKPNPPLLLGGECAGRIVAVGE-GVNGLVVGQPVIALSAGA 1042
            :|   ||  ||  ||  |  :  |  :          ||        |     : ||:   :   |   |  |        |||
Db    1333 ISRKDVNGLARGHLVDGI------PLCTPSFYADIAMQVGQYSMQRLRAGHP----GAGA 1382

Qy    1043 FATHVTTSAALV----LPR---PQAL-SAIEAAAMPVAYLTAWYALDRIAR-LQPGERVL 1093
                |   |     :|        :|       | |  :  :          |  |     |   :  |          :
Db    1383 IDGLVDVSDMVVDKALVPHGKGPQLLRTTLTMEWPPKAAATTRSAKVKFATYFADGKLDT 1442

Qy    1094 IHAA-----TGGVGLAAV-----QWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRS- 1142
             ||:        |       |  ::        ::   |:   ::|:         |        |   : :|
Db    1443 EHASCTVRFTSDAQLKSLRRSVSEYKTHI-RQLHDGHAKGQFMRYNRKTGYKLMSSMARF 1501

Qy    1143 -------DRFVADVRAWTGGEGVDVVLNSLSGEL-------------------------I 1170
                         |   |  :          |||   |  |   |                                |
Db    1502 NPDYMLLDYLVLNEAENEAASGVDFSLGSSEGTFAAHPAHVDAITQVAGFAMNANDNVDI 1561

Qy    1171 DK---------SFNLL------RSHGRFVELGKRDCYADNQL-----------GLRPFLR 1204
            :|                 || :           :|:   :  ::|:       :| |                    :   | |
Db    1562 EKQVYVNHGWDSFQIYQPLDNSKSYQVYTKMGQA---KENDLVHGDVVVLDGEQIVAFFR 1618

Qy    1205 NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFT----PPPIATLPIARVADAFRSM 1260
             |:      |      :  :        |:       :|         |  |         |||   |:|:|:        |          :
Db    1619 GLTLRSVPRGALRVVLQTTVKKADRQL------GFKTMPSPPPPTTTMPISPYKPANTQV 1672

Qy    1261 ------AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAA-- 1312
                      |:|  |          :       |:          | |:         |               :||||:
Db    1673 SSQAIPAEATH------SHTPPQ-----PKHSPV-PET---------AGSAPAAKGVGVS 1711

Qy    1313 ---LEAFLRTQVSQVLRTPEIKVGAEAL-----FTRLGMDSLMAVELRNRIEASLKLKLS 1364
                   |:| :|  ||:            |   :   | |           |    :|:|||  :: : :|        | | |
Db    1712 NEKLDAVMRV-VSE-----ESGIALEELTDDSNFADMGIDSLSSMVIGSRFREDLGLDLG 1765

Qy    1365 TTF 1367
             |
Db    1766 PEF 1768
```

RESULT 11

```
FAS_RAT
ID   FAS_RAT        STANDARD;      PRT;  2505 AA.
AC   P12785; O09187; O09190; Q64717;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89240686; PubMed=2717611;
RA   Amy C.M., Witkowski A., Naggert J., Williams B., Randhawa Z.,
RA   Smith S.;
RT   "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT   fatty acid synthase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=93075999; PubMed=1339331;
RA   Beck K.F., Schreglmann R., Stathopulos I., Klein H., Hoch J.,
RA   Schweizer M.;
RT   "The fatty acid synthase (FAS) gene and its promoter in Rattus
RT   norvegicus.";
RL   DNA Seq. 2:359-386(1992).
RN   [3]
RP   SEQUENCE OF 75-2505 FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
RX   MEDLINE=89128431; PubMed=2915923;
RA   Schweizer M., Takabeyashi K., Beck K.F., Schreglmann R.;
RT   "Rat mammary gland fatty acid synthase: localization of the
RT   constituent domains and two functional polyadenylation/termination
RT   signals in the cDNA.";
RL   Nucleic Acids Res. 17:567-586(1989).
RN   [4]
RP   SEQUENCE OF 2085-2505 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=88087240; PubMed=2891707;
RA   Naggert J., Witkowski A., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the thioesterase
RT   domain of the rat fatty acid synthetase.";
RL   J. Biol. Chem. 263:1146-1150(1988).
RN   [5]
RP   SEQUENCE OF 1921-2324 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=87246646; PubMed=3109907;
RA   Witlowski A., Naggert J., Mikkelsen J., Smith S.;
RT   "Molecular cloning and sequencing of a cDNA encoding the acyl carrier
RT   protein and its flanking domains in the mammalian fatty acid
RT   synthetase.";
RL   Eur. J. Biochem. 165:601-606(1987).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
```

```
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M76767; AAA57219.1; -.
DR   EMBL; X62888; CAA44679.1; -.
DR   EMBL; X62889; CAA44680.1; -.
DR   EMBL; X13415; CAA31780.1; -.
DR   EMBL; X13527; CAA31882.1; -.
DR   EMBL; J03514; AAA41144.1; -.
DR   PIR; A30313; XYRTFA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR000051; SAM_bind.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00107; ADH_zinc_N; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW   Pyridoxal phosphate.
FT   DOMAIN        1    413       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      429    817       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1629   1857       ENOYL REDUCTASE.
FT   DOMAIN     1858   2113       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2118   2174       ACYL CARRIER (ACP).
FT   DOMAIN     2202   2505       THIOESTERASE.
FT   ACT_SITE    161    161       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    581    581       MALONYLTRANSFERASE (BY SIMILARITY).
FT   NP_BIND    1665   1682       NADP (ER).
FT   BINDING    1698   1698       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2151   2151       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2302   2302       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2475   2475       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE    878    878       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                SIMILARITY).
FT   CONFLICT    871    871       S -> P (IN REF. 3).
FT   CONFLICT   1967   1968       MV -> IL (IN REF. 5).
FT   CONFLICT   2085   2085       C -> P (IN REF. 4).
FT   CONFLICT   2106   2106       A -> V (IN REF. 1 AND 5).
FT   CONFLICT   2296   2296       Y -> H (IN REF. 1 AND 5).
SQ   SEQUENCE   2505 AA;  272647 MW;  5810EC13D37F3114 CRC64;

  Query Match             15.3%;  Score 1100;  DB 1;  Length 2505;
  Best Local Similarity   22.7%;  Pred. No. 3.4e-50;
  Matches  470;  Conservative  221;  Mismatches  586;  Indels  792;  Gaps   56;

Qy         12 DPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKTPVT 71
              : : | | | :||    :|. ||  | |  | |       || |  :  |
Db          2 EEVVIAGMSGKLPESE-NLQEFWANLIGGVDMV-TDDDRRWKAGLYGLP----------K 49

Qy         72 RASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGVFI 131
              |:  | |:: ||||||: |::|  |||  ||||| :||: :   | |::| || |||::
Db         50 RSGKLKDLSKFDASFFGVHPKQAHTMDPQLRLLLEVSYEAIVDGGINPASLRGTNTGVWV 109

Qy        132 GIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSSLVA 191
              |:  ||   || :   :  :  :  :|    ::: | |:|:   :|| :|:||| ||||:|
Db        110 GVSGSEASEALSRDPET--LLGYSMVGCQRAMMANRLSFFFDFKGPSIALDTACSSSLLA 167

Qy        192 VHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGRGEG 251
               :  | |::||||  |:  ||::|:| |:|  |      |:  || |::|     :|: |  |
Db        168 LQNAYQAIRSGECPAAIVGGINLLLKPNTSVQFMKLGMLSPDGTCRSFDDSGNGYCRAEA 227

Qy        252 CAVVVLKRLSGARADGDRILAVIRGSAINHDGA-SSGLTVPNGSSQEIVLKRALADAGCA 310
               |:| :  | ||     |:  |  |     | ||   |:| |:| :|| :::       | |
Db        228 VVAVLLTKKSLAR----RVYATILNAGTNTDGCKEQGVTFPSGEAQEQLIRSLYQPGGVA 283

Qy        311 ASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVA----TPLLIGSVKTNLGHPEYASG 366
                |: |:||||||| :|||  | ||   |: |  :       :||||| |:|:|||| |||
Db        284 PESLEYIEAHGTGTKVGDP---QELN---GITRSLCAFRQSPLLIGSTKSNMGHPEPASG 337

Qy        367 ITGLLKVVLSLQHGQIPAHLHAQALNPRI-SWGDLRLTVTRARTPWPDWNTPRR---AGV 422
               :  | ||:|||::|     :||    || |  :  | || |          |  | |   |:
Db        338 LAALTKVLLSLENGVWAPNLHFHNPNPEIPALLDGRLQVV-------DRPLPVRGGIVGI 390

Qy        423 SSFGMSGTNAHVVL----EEAPAATCTPPAPERP-AELLVLSARTASALDAQAARLRDHL 477
              :|||  | | ||:|    ::|||      |||       ||  | || |:      : | |
Db        391 NSFGFGGANVHVILQPNTQQAPA-----PAPHAALPHLLHASGRTMEAVQGLLEQGRQHS 445

Qy        478 ETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSS 537
              :          |:||      :| :  :|   |:    |          :|      |: : |
Db        446 Q--------DLAF------VSMLNDIAATPTAAMPFRGYTVLGVEGHVQ--EVQQVPASQ 489

Qy        538 RGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVD 597
              |  | |: :| | |  |  |  |  |  :  :|||:     |  :  |  :||  |  :  |
Db        490 R-PLWFICSGMGTQWRGMGLSLMRL-DSFRESI-----LRSDEALKPLGVKVSDLLLSTD 542

Qy        598 AALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFL 657
                 |     :  :|    : ||   |   | |::|: :  |||:||:       |   |  :||
Db        543 EHTFDDIVHSFVSLTAIQIALIDLLTSMGLKPDGIIGHSLGEVACGYADGCLSQREAVLA 602

Qy        658 VAARGRLMQ--ALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPV 715
                 ||: ::   ||| |:| ::     :     |    |   |  |:  |  | |  |:|   |
Db        603 AYWRGQCIKDANLPA-GSMAAVGLSWEECKQRCPP---GVVPACHNSEDTVTISGPQAAV 658

Qy        716 HAIAAAMAARGARTKALHVSH-AFHSPLM---AP-MLEAFGRV--------AESVSYRRP 762
              :      :     |  :    ||||  |     ||  :|:|  :|          |  :|  |
Db        659 NEFVEQLKQEGVFAKEVRTGGLAFHSYFMEGIAPTLLQALKKVIREPRPRSARWLSTSIP 718

Qy        763 SIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTL--- 819
                |:|:              ||  | |  |  :  :  |          :|:  |  :  :|
Db        719 EAQWQSSLA------RTSSAEYNVNNLVSPVLFQEAL--WHVPEHAVVLEIAPHALLQAV 770

Qy        820 --LGLVPACMPDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVP- 876
                 |: |:|     :    |  :|     |  || :   |  ::  |||    ||
Db        771 LKRGVKPSC-----TIIPLMKRDHKDNLEFFLTNLGKVHLTGIDINPNALFPPVEFPVPR 825

Qy        877 -----LPTYPWQRERYWIDTKADDAARG--------------------------DRRA- 903
                    |   |   :  |     |:|   |                            ||
Db        826 GTPLISPHIKWDHSQTWDIPVAEDFPNGSSSSSATVYNIDASSESSDHYLVDHCIDGRVL 885

Qy        904 -PGAGH-----------------------------------------------------D 909
                 || |:                                                     :
Db        886 FPGTGYLYLVWKTLARSLSLSLEETPVVFENVTFHQATILPRTGTVPLEVRLLEASHAFE 945

Qy        910 EVEEGGAVRGG------DRRSARLDHP----PPES------------------------ 934
                :  |  :        |    |   |||     | ||
Db        946 VSDSGNLIVSGKVYQWEDPDSKLFDHPEVPIPAESESVSRLTQGEVYKELRLRGYDYGPH 1005

Qy        935 ------------------------------------------------------------ 934

Db       1006 FQGVYEATLEGEQGKLLWKDNWVTFMDTMLQISILGFSKQSLQLPTRVTAIYIDPATHLQ 1065

Qy        935 ------------------------------------------------------------ 934

Db       1066 KVYMLEGDTQVADVTTSRCLGVTVSGGVYISRLQTTATSRRQQEQLVPTLEKFVFTPHVE 1125
```

```
Qy        935 ------------------------------GRREKVEAAGDRP--------------- 947
                                            | :  |   | |
Db       1126 PECLSESAILQKELQLCKGLAKALQTKATQQGLKMTVPGLEDLPQHGLPRLLAAACQLQL 1185

Qy        948 ---FRLEIDEP------------------------------------------------- 955
                 :||: |
Db       1186 NGNLQLELGEVLARERLLLPEDPLISGLLNSQALKACIDTALENLSTLKMKVVEVLAGEG 1245

Qy        956 ---------------------------------------------------------GVL 958
                                                                       | |
Db       1246 HLYSHISALLNTQPMLQLEYTATDRHPQALKDVQTKLQQHDVAQGQWDPSGPAPTNLGAL 1305

Qy        959 D-------------------------------HLVLR--------------------- 964
              |                                | ||:
Db       1306 DLVVCNCALATLGDPALALDNMVAALKDGGFLLMHTVLKGHALGETLACLPSEVQPGPSF 1365

Qy        965 --------VTERRA---------------------------------------------- 970
                      :  |:|
Db       1366 LSQEEWESLFSRKALHLVGLKKSFYGTALFLCRRLSPQDKPIFLPVEDTSFQWVDSLKSI 1425

Qy        971 --------------------------------------------------------PGLGE 975
                                                                       ||  |
Db       1426 LATSSSQPVWLTAMNCPTSGVVGLVNCLRKEPGGHRIRCILLSNLSSTSHVPKLDPGSSE 1485

Qy        976 VEIAVDA----------------------------------------------------- 982
              ::  :::
Db       1486 LQKVLESDLVMNVYRDGAWGAFRHFQLEQDKPEEQTAHAFVNVLTRGDLASIRWVSSPLK 1545

Qy        983 -----------------AGLSFNDVQLALG-MVPDDLPGK-PNPPLLLGGECAGRIVAVG 1023
                               | |:| |: || | : || :|||  :   :|| | :||    |
Db       1546 HMQPPSSSGAQLCTVYYASLNFRDIMLATGKLSPDAIPGKWASRDCMLGMEFSGR-DKCG 1604

Qy       1024 EGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRI 1083
                 | |||          |   || |  |      :   | : :  |||::|| | ||:|:|
Db       1605 RRVMGLV--------PAEGLATSVLLSPDFLWDVPSSWTLEEAASVPVVYTTAYYSLVVR 1656

Qy       1084 ARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLES----LGVRYVSD 1139
                |:| || |||: :|||  ||:   |   :|   |   | |: ||||||::     |       ::
Db       1657 GRIQHGETVLIHSGSGGVGQAAISIALSLGCRVFTTVGSAEKRAYLQARFPQLDDTSFAN 1716

Qy       1140 SRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGL 1199
               ||    |    |  |||:|||:|||||: |  :  |    |  |||:|:|:|| |    :: ||:
Db       1717 SRDTSFEQHVLLHTGGKGVDLVLNSLAEEKLQASVRCLAQHGRFLEIGKFDLSNNHPLGM 1776

Qy       1200 RPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRS 1259
                ||:|::|   : |   :              | : | |    |  ||   |     | |:| ||||
Db       1777 AIFLKNVTFHGILLDALFEGANDSWREVAELLKAGIRDGVVKPLKCTVFPKAQVEDAFRY 1836

Qy       1260 MAQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRT 1319
              |||  :|:|:||::: : : |  :  :|        || |:            |||        :::: |
Db       1837 MAQGKHIGKVLVQVREEEPEAMLP---GAQPT---------LISAISKTFCPEHKSYIIT 1884

Qy       1320 --------QVSQ--VLRTPE---------IKVGAEAL----FTRLGMDSLMAVELRNRIE 1356
                      ::::  |||   :          |: |  :|     :  | |:   |::    :  :|
Db       1885 GGLGGFGLELARWLVLRGAQRLVLTSRSGIRTGYQAKHVREWRRQGIHVLVSTSNVSSLE 1944

Qy       1357 ASLKLKLSTTFLSTSPNIALLAQNLLDAL 1385
               :  |    | |      :  ||  | ||:
Db       1945 GARALIAEATKLGPVGGVFNLAMVLRDAM 1973
```

RESULT 12
STCA_EMENI

```
ID   STCA_EMENI     STANDARD;      PRT;   2181 AA.
AC   Q12397;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative sterigmatocystin biosynthesis polyketide synthase (PKS).
GN   STCA OR PKSST.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=162425;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 26;
RX   MEDLINE=96202293; PubMed=8643646;
RA   Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA   Keller N.P., Adams T.H., Leonard T.J.;
RT   "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT   cluster in Aspergillus nidulans.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FGSC 4;
RX   MEDLINE=95370159; PubMed=7642507;
RA   Yu J.-H., Leonard T.J.;
RT   "Sterigmatocystin biosynthesis in Aspergillus nidulans requires a
RT   novel type I polyketide synthase.";
RL   J. Bacteriol. 177:4792-4800(1995).
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC       STERIGMATOCYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC       (POTENTIAL).
CC   -!- PATHWAY: Sterigmatocystin biosynthesis; first step.
CC   -!- SIMILARITY: Contains 2 acyl carrier domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U34740; AAC49191.1; -.
DR   EMBL; L39121; AAA81586.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
KW   Transferase; Acyltransferase; Phosphopantetheine; Repeat;
KW   Multifunctional enzyme.
FT   DOMAIN      383    814       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      884   1209       ACYL/MALONYL TRANSFERASES.
FT   DOMAIN     1706   1777       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1830   1901       ACYL CARRIER (ACP) 2.
FT   DOMAIN        ?   2181       THIOESTERASE.
FT   ACT_SITE    552    552       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    978    978       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1738   1738       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1862   1862       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2028   2028       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2181 AA;  238831 MW;  5A3E5712AA9AD942 CRC64;

  Query Match             15.0%;  Score 1083;  DB 1;  Length 2181;
  Best Local Similarity   25.9%;  Pred. No. 2.3e-49;
  Matches  402;  Conservative  235;  Mismatches  637;  Indels  280;  Gaps   53;

Qy          7 ERAAEDP------IAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
              || :  |      :|||  | |            || ||    |     || |||   ||
Db        370 ERYSHRPGSDRGKLAIVSMSGRFPEAP-STDSFWDLLYKGLDVCKEVPLRRWDVKTHVDP 428

Qy         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
```

```
                  |  |          :|       ||  || |||:|| :|||| |: |    :||:|   | |
Db       429 SGKARNKGATRWGCWLDFAGEFDPRFFSISPKEAPQMDPAQRMALMSTYEAMERGGIVPD 488

Qy       121 ALVGTE---TGVFIGIGPSEYEAALPQATASAEIDAH----GGLGTMPSVGAGRISYALG 173
                    |:      ||| |: :::      :   :   || :     |  | :|     |||::
Db       489 TTPSTQRNRIGVFHGVTSNDW----METNTAQNIDTYFITGGNRGFIP----GRINFCFE 540

Qy       174 LRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARD 233
                  ||  : |||  ||||  |:||||  ||   |:|  ||:|||  ::: :|     | |   |
Db       541 FSGPSYSNDTACSSSLAAIHLACNSLWRGDCDTAVAGGTNMIFTPDGHTGLDKGFFLSRT 600

Qy       234 GRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNG 293
              | ||||    |||:  | ||     |  :|||    | |:  | |||  |        ||     |  :| |
Db       601 GNCKAFDDAADGYCRAEGVGTVFIKRLEDALAENDPILATILDIKTNHSAMSDSMTRPFK 660

Qy       294 SSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYG---LGRDVATPLL 350
              :|    :     |:  ||  :    : |:|  |||||  :||   :|:::: :::          |      ||
Db       661 PAQIDNMSALLSTAGISPLDLSYIEMHGTGTQVGDAVEMESVLSLFAPDETFRPRDKPLY 720

Qy       351 IGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNP--RI--SWGDL---RLT 403
             :|| |  |:|| |     ||:| |:||:| :::    || |       :  |   ||  :: ||       :
Db       721 VGSAKANIGHGEGVSGVTSLIKVLLMMKNDTIPPHC---GIKPGSRINRNYPDLPARNVH 777

Qy       404 VTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTA 463
             :        |||    :||||   :::|   | |  |:|:|||         |||      ::
Db       778 IAFEPKPWPRTDTPRRVLINNFSAAGGNTAVLVEDAPV-----------PELGEIAGSFG 826

Qy       464 SALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAV-AATSREGLRAALDAAAQ 522
              |    |::  |                 :::        |    ||:::   |  :  |  ||      ||  |:
Db       827 EAAGNQSSEL---------------SYTTTARRWHHPHRVSITGANTMEILRNVESAIAR 871

Qy       523 GQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFR---EAFDLCVRLFNQ 579
              |    |   |     ::: |:    :|||:|   |||   ||: :    ||     | ||  .  | :
Db       872 GH---GVNR---PATKPKIVIACSGQGSQYTGMGWQLYNSYPTFRSDLERFDQLARSYGF 925

Qy       580 ELDRPLREVMWAEPA--SVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSI 637
                       ||   ::|    |::   |               | ||  :  |  ||   |   |:|::|  |  |||:
Db       926 P---SFLEVYTSKPVGDSMEDLL---PVIVQLALVSLEMALGNLLGSFGLKPSAVIGHSL 979

Qy       638 GELVAACVAGVFSLEDAVFLVAARGRLMQALPAGG--AMVSIEAPEADVAAAVAPHAASV 695
              ||   |   ::||   |   |  ::||   |  |:|           |  ||:::   |      :|      ::
Db       980 GEYAALYISGVLSAADTLYLVGMRAKLLQERCQRGTHAMLAVRASPVTLCEVLA--ESNC 1037

Qy       696 SIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAE 755
              :|    | |:   |::|    :    |  :    :::|   | :      |  :   ||||   :  |:||  |  ||
Db      1038 EVACHNGPNDTVLSGPLKEVMNLQNSLSATGIKGTLLKLPFAFHSAQVQPILEEFKNVAR 1097

Qy       756 SVSYRRPSIVLVSNLSGKACTDE-VSSPGYWVRHAREVVRFA---DGVKALHAAGAGTFV 811
              |:: :| |  ::|  |    |         ::        |  |    ||  || |:          :    :           ||
Db      1098 GVTFHKPQIPVLSPLLVKVIDEKGTVDPVYLARHCREPVKMVSVLEHARDQHIITDRTIV 1157

Qy       812 -EVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGL--- 867
              :||||: : |::    :           : |  :          |      ::      ||  |:: |   ::|
Db      1158 IDVGPKALMAGMIKTTLDKDTSSALPTLGPSLDVWKSLTNILGTLYSRGLDINWVAYHEP 1217

Qy       868 FPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEGGAVRGGDRRSARL 927
              |  |    :  :  ||:|  |    :  :|:|     |  :                      |:          |     |     :  ::
Db      1218 FGSAKKVIELPSYGWDLKDYFIPYKGEWCLH---------RHEIRCSCATPGKETATSDY 1268

Qy       928 DHPPPESGRREKVEAAGDRPPRLEIDEPG----VLDHLVLRVTERRAPGLGEVEIAVDAA 983
                 |            |  |||    :    |  :    :          |   |||  |   :   ||   |
Db      1269 QLPSDE-------QVAAKRPSKQDESKEAYPEIVATTVHRVVEEKTEPLGAT--LVVET 1319

Qy       984 GLSFNDV-QLALGMVPDDLP-----GKPNPPLLLGGECAGRIVAVGEG-VNGLV-VGQPV 1035
              :|   ||  |:||  :  |  :|           :  |  :|          |:  |    |  ::|:|  |   |
Db      1320 DISRPDVNQIAQGHLVDGIPLCTPSVYADIALHVGRYSMNRLRASHPGAMDGVVDVADMV 1379

Qy      1036 I---ALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERV 1092
              |      :   |      :  |:   :    |  |:|  :     :|   :     |      |     ||
Db      1380 IDKALIPHGKSPQLLRTTLTMTWP-PKAAATTRSAKIKFATYFADGKLD----------- 1427

Qy      1093 LIHAATGGVGLAAVQWAQHVGAEVHATA---------GTPEKRAYLESLG--------VR 1135
                                   :|    |  |:                 ||  :  ::  ||                   :|

Db      1428 ----------------TEHATCTVRFTSEAQLKSLQKKVPEYQERIKKLGEGLRQGQFIR 1471

Qy      1136 YVSD------SRSDRFVADVRAWTGGEGVDVVLNSLSGELID-KSFNLLRSHGRFVELGK 1188
              | :          |     |    |  |  :                :: ||     |  :  ·  |:   :|  |  |
Db      1472 YTTKSGYKLMSSMASFHRDYKLLN-----HLILNEADNEAVSTMDFSAAKSEGTF---AA 1523

Qy      1189 RDCYAD--NQLG--------------------------LRPFLRNLSFSLV-----DLR 1214
                  |  |      |:|                                     :|  ::   :: :               |  :
Db      1524 HPAYVDAITQVGGFAMNANDNTDIQQEVFVNHGWTSFQVYQPLVKGKTYEVYVRMTEDEK 1583

Qy      1215 GMMLERPARV---RALLEELLGL-IAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLV 1270
              | ::       |        |::   ||  ::   :  :   :   :  ||        |      |    |
Db      1584 GDLVHGDTIVLYGDAVVAFFKGLSVSLSHLSETQLTSWYVRRVPRRGLRMVLQQASDKAA 1643

Qy      1271 LTLGDPE-VQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPE 1329
                 |: :  |: :  |  |                  |  :     ::|       |:    |:  |:   |        |
Db      1644 RLHGNQQAVKTQAPQRAA----------LKQKPQSSPTQPHASKVAYSRSATSPTAGKPV 1693

Qy      1330 I----------------------------KVGAEALFTRLGMDSLMAVELRNRIEASLK 1360
              :                                   :: |:    |    :|:|||  :: :  :|:       |
Db      1694 VAARDLSREGDDKFKAVLSVISEESGVALGELTADTNFADIGIDSLSSMVIGSRLREDLG 1753

Qy      1361 LKLST---------------TFLSTSPNIALLAQNLLDALATALSLERVAAENL 1399
              |:|                       |  ||  |     |:     |    |  |          |  |  |  |
Db      1754 LELGAEFSLFIDCPTVRSLKTLLSGS---AVSVNNDKDELEPGQEAETAAPEQL 1804
```

```
RESULT 13
FAS_HUMAN
ID   FAS_HUMAN      STANDARD;      PRT;   2504 AA.
AC   P49327;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN OR FAS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96004605; PubMed=7567999;
RA   Jayakumar A., Tai M.-H., Huang W.-Y., Al-Feel W., Hsu M.,
RA   Abu-Elheiga L., Chirala S.S., Wakil S.J.;
RT   "Human fatty acid synthase: properties and molecular cloning.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:8695-8699(1995).
RN   [2]
RP   SEQUENCE OF 753-758 AND 1285-1297.
RX   MEDLINE=94294385; PubMed=8022791;
RA   Kuhajda F.P., Jenner K., Wood F.D., Hennigar R.A., Jacobs L.B.,
RA   Dick J.D., Pasternack G.R.;
RT   "Fatty acid synthesis: a potential selective target for
RT   antineoplastic therapy.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:6379-6383(1994).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC       LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC       THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC       ACYL CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
```

```
CC        = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC    -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC        2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC    -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC        carrier protein] + oleate.
CC    -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC    -!- TISSUE SPECIFICITY: PROMINENT EXPRESSION IN BRAIN, LUNG, AND
CC        LIVER.
CC    -!- MISCELLANEOUS: THE RELATIVELY LOW BETA-KETOACYL SYNTHASE ACTIVITY
CC        MAY BE ATTRIBUTABLE TO THE LOW 4'-PHOSPHOPANTETHEINE CONTENT OF
CC        THE PROTEIN.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC    the European Bioinformatics Institute.  There are no  restrictions on  its
CC    use  by  non-profit  institutions as long  as its content  is  in  no  way
CC    modified and this statement is not removed.  Usage  by  and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------------
DR    EMBL; U26644; AAC50259.1; -.
DR    PIR; A57788; A57788.
DR    Genew; HGNC:3594; FASN.
DR    MIM; 600212; -.
DR    InterPro; IPR001227; Ac_transferase.
DR    InterPro; IPR002085; Adh_zn_family.
DR    InterPro; IPR000794; Ketoacyl-synt.
DR    InterPro; IPR006163; Pp_bind.
DR    InterPro; IPR006162; Ppantne_attach.
DR    InterPro; IPR000051; SAM_bind.
DR    InterPro; IPR001031; Thioesterase.
DR    Pfam; PF00698; Acyl_transf; 1.
DR    Pfam; PF00107; ADH_zinc_N; 1.
DR    Pfam; PF00109; ketoacyl-synt; 1.
DR    Pfam; PF02801; ketoacyl-synt_C; 1.
DR    Pfam; PF00550; pp-binding; 1.
DR    Pfam; PF00975; Thioesterase; 1.
DR    PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR    PROSITE; PS50075; ACP_DOMAIN; 1.
DR    PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
KW    Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW    Hydrolase; Oxidoreductase; Transferase; Lyase; NADP;
KW    Pyridoxal phosphate.
FT    DOMAIN        1    413       BETA-KETOACYL SYNTHASE.
FT    DOMAIN      428    815       ACYL AND MALONYL TRANSFERASES.
FT    DOMAIN     1630   1857       ENOYL REDUCTASE.
FT    DOMAIN     1858   2113       BETA-KETOACYL REDUCTASE.
FT    DOMAIN     2118   2174       ACYL CARRIER (ACP).
FT    DOMAIN     2202   2504       THIOESTERASE.
FT    ACT_SITE    161    161       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT    ACT_SITE    580    580       MALONYLTRANSFERASE (BY SIMILARITY).
FT    ACT_SITE    876    876       BETA-HYDROXYACYL DEHYDRATASE (BY
FT                                 SIMILARITY).
FT    NP_BIND    1666   1683       NADP (ER).
FT    BINDING    1699   1699       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT    NP_BIND    1880   1895       NADP (KR).
FT    BINDING    2151   2151       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT    ACT_SITE   2302   2302       THIOESTERASE (BY SIMILARITY).
FT    ACT_SITE   2475   2475       THIOESTERASE (BY SIMILARITY).
SQ    SEQUENCE   2504 AA;  273100 MW;  8AAF9809B2338DFA CRC64;

  Query Match             14.4%;  Score 1039;  DB 1;  Length 2504;
  Best Local Similarity   23.5%;  Pred. No. 5.6e-47;
  Matches  477;  Conservative  208;  Mismatches  582;  Indels  766;  Gaps   68;

Qy         12 DPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKTPVT 71
                : :  |  |     :||     :|   ||  |  |  |  |        || |   :   |
Db          2 EEVVIAGMFGKLPESE-NLQEFWDNLIGGVDMV-TDDDRRWKAGLYGLP---------R 49

Qy         72 RASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGVFI 131
              |:  | |::  ||||||: |::|  |||   |||||  :||: :  | | :| || |||::
Db         50 RSGKLKDLSRFDASFFGVHPKQAHTMDPQLRLLLEATYEAIVDGGINPDSLRGTHTGVWV 109

Qy        132 GIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSSLVA 191
              |:  ||  ||   :  :   :  :  :|   ::  | |:|:  ||| :|:||| ||||:|
Db        110 GVSGSETSEALSRDPET--LVGYSMVGCQRAMMANRLSFFFDFRGPSIALDTACSSSLMA 167

Qy        192 VHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGRGEG 251
               :  | |:: ||:|  |: ||::::| |:|  |     |:  :| ||||      :|: |  ||
Db        168 LQNAYQAIHSGQCPAAIVGGINVLLKPNTSVQFLRLGMLSPEGTCKAFDTAGNGYCRSEG 227

Qy        252 CAVVVLKRLSGARADGDRILAVIRGSAINHDG-ASSGLTVPNGSSQEIVLKRALADAGCA 310
                 |:| : | ||      ::     |      | ||      |:|  |      ||  ::       || |
Db        228 VVAVLLTKKSLAR----KVYTTILNKGTNTDGFKEQGVTFPQ-DIQEQPIRSLYQSAGVA 282

Qy        311 ASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGL 370
                 |    |:||||  || :|||  |        :           |     ||||||  |:|:||||  |||:  |
Db        283 PESFEYIEAHGPGTKVGDPQERNGITRALCATRQ--EPLLIGSTKSNMGHPEPASGLDAL 340

Qy        371 LKVVLSLQHGQIPAHLHAQALNPRI-SWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSG 429
               ||:|||:||       :||   :  || |  :    | ||   |         |           |       |::|||  |
Db        341 AKVLLSLEHGLWAPNLHFHSPNPEIPALLDGRLQVVDQPLPVRGGN----VGINSFGFGG 396

Qy        430 TNAHVVL----EEAPAATCTPPAPERP-AELLVLSARTASALDAQAAR-LRDHLETYPSQ 483
               :| |::|        : |||        |||          ||   ||   |:              :  ||
Db        397 SNMHIILRPNTQSAPA-----PAPHATLPRLLRASGRTPEAVQKLLEQGLRHSQGLAFLS 451

Qy        484 CLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAF 543
               | |:|      |    :||  |              |          |       |:|        |: :       |   |
Db        452 MLNDIA---AVPATAMPFR---------GY-----AVLGGETRWPRVQQVPAGER-PLWF 493

Qy        544 LFTGQGAQTLGMGRGLYDVWSAFREAF---DLCVRLFNQELDRPLREVMWAEPASVDAAL 600
               : :|  |  |   |||    |   :      ||::       |     |      |     :: :  |              |  |  :
Db        494 ICSGMGTQWRGMGLSLMRL-DRFRDSILRSDEAVNRFGLKVSQLL--------LSTDEST 544

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
                 |     :     :|     :      |     |      |  |  :  :  |||:||  ::         |     |  :||
Db        545 FDDIVHSFVSLTAIQIGLIDLLSCMGPEADGIVGHSLGEWLSV-RDGCLSQEEAVLAAYW 603

Qy        661 RGRLMQALP-AGGAMVSI----EAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPV 715
               ||: ::       |  |||    :           |   :                 |  |:   |   |  |:|      ||
Db        604 RGQCIKEAPLPAGAMAAVGLSWEECKQRCPPAVVP-------ACHNSKDTVTISGPQAPV 656

Qy        716 HAIAAAMAARGARTKALHV-SHAFHSPLM----APMLEAFGRV---------------- 753
                       :        |       |    :          ||||      |                 |:|:        :|
Db        657 FEFVEQLRKEGVFAKEVRTGGMAFHSYFMEAIAPPLLQELKKVIREPKPRSARWLSTSIP 716

Qy        754 ----------AESVSY---------------------------------------RRPS 763
                                        |     |                                                                        |:||
Db        717 EAQWHSSLARTSSAEYNVNNLVSPVLFQEALWHVPEHAVVLEIAPTPCPQAVLKRVRKPS 776

Qy        764 IVLVS-------------------NLSG------------------------------ 772
                 ::                                         :|||
Db        777 CTIIPRMKKDHRDNLEFFLAGIGRLHLSGIDANPNALFPPVESPAPRGTPLISPLIKWDH 836

Qy        773 -----------------------KACTDEVSSP-----------------GY------- 784
                                                              ||    ||                                         ||
Db        837 SLAWDAPAAEDFPNGSGSPSATIYTCTPSSESPDRYLVDHTIDGRVLFPATGYLSIVWKT 896

Qy        785 ----WVRHAREVVRFAD----------------GVKALHAAGA------GTFVEVG--- 814
                         |        :     |  |  |                                  |:  |  |  |  ||            |    |    |
Db        897 LARAWAGLEQLPVVFEDVVQHQATILPKTGTVSLEVRLLEATGAFEVSENGNLVVSGKVY 956

Qy        815 --------------------PKSTLL---------------------GLVPACMPD--- 829
                                                    |:|  |                                          |::  |  :
Db        957 QWDDPDPRLFDHPESPHPNSPRSPLFLAQAEVYKELRLRGYDYGPHFQGILEASLEGDSG 1016

Qy        830 ---------------ARPALLASSRAG----------RDEPAT-------------VLEA 851
                                        :  ::|  |::  |                              :|||                             |  :
Db       1017 RLLWKDNWVSFMDTMLQMSILGSAKHGLYLPTRVTAIHIDPATHRQKLYTLQDKAQVADV 1076

Qy        852 LGGLW---AVGGLVSWAGLFPSGGRR------VPL------------------------ 877
               :     |  |  |   :||                       ||:
Db       1077 VVSRWPRVTVAGGVHISGLHTESAPRRHEEQQVPILEKFCFTPHTEEGCLSEHAALEEEL 1136

Qy        878 ------------------------------PTYPWQRE--------------------- 885
```

```
                                                |  | |:|
Db      1137 QLCKGLVEALETKVTQQGLKMVVPDWTGPRSPRDPSQQELPRLLSAACRLQLNGNLQLEL 1196

Qy       886 --------------------------RYWIDTKADD------------AARGD--RRAPG 905
                                       :  :||  ::            |  |    ||
Db      1197 AQVLAQERPKLPEDPLLSGLLDSPALKACLDTAVENMPSLKMKVVEVLAGHGHLYSRIPG 1256

Qy       906 ----------------------AGHDEVEEGGAVRG---------------------- 919
                                   |  |:::   :|
Db      1257 LLSPHPLLQLSYTATDRHPQALEAAQAELQQHDVAQGQWDPADPAPSALGSADLLVCNCA 1316

Qy       920 ----GDRRSARLD-------------------HP----------PPESGR---------- 936
                 || ||  :                    ||           |: |:
Db      1317 VAALGDPASALSNMVAALREGGFLLLHTLLRGHPRDIVAFLTSTEPQYGQGILSQDAWES 1376

Qy       937 ---REKVEAAG-------------------DRPFRLEID--------------------- 953
                |  :  |                    |  |  | :|
Db      1377 LFSRVSLRLVGLKKSFYGATLFLCRRPTPQDSPIFLPVDDTSFRWVESLKGILADEDSSR 1436

Qy       954 ----------------------EPG-----------------------------VLDH 960
                                   |||                              :
Db      1437 PVWLKAINCATSGVVGLVNCLRREPGGTVRCVLLSNLSSTSHVPEVDPGSAELQKVLQGD 1496

Qy       961 LVLRVTERRAPGL--------------------------------------------GE 975
             ||: |    | |:                                            |
Db      1497 LVMNVYRDGAWGVFRHFLLEDKPEEPTAHAFVSTLTRGDLSSIRWVCSSLRHAQPTCPGA 1556

Qy       976 VEIAVDAAGLSFNDVQLALG-MVPDDLPGK-PNPPLLGGECAGRIVAVGEGVNGLVVGQ 1033
                  |  |:|  |: ||  |  : || :|||   :      |||  |  :||   |  |: |  |||
Db      1557 QLCTVYYASLNFRDIMLATGKLSPDAIPGKWTSQDSLLGMEFSGR-DASGKRVMGLV--- 1612

Qy      1034 PVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVL 1093
                     |     ||  |   |    :     |     :   |||::|| | ||:|||     |::||| :|
Db      1613 -----PAKGLATSVLLSPDFLWDVPSNWTLEEAASVPVVYSTAYYALVVRGRVRPGETLL 1667

Qy      1094 IHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLES----LGVRYVSDSRSDRFVADV 1149
             ||: :||||  ||:   |   :|   |   |  | |: ||||||::     |      ::||   |    |
Db      1668 IHSGSGGVGQAAIAIALSLGCRVFTTVGSAEKRAYLQARFPQLDSTSFANSRDTSFEQHV 1727

Qy      1150 RAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLRNLSFS 1209
                  |||:|||:|||||: |  :    |          :||||:|:||  |      ::  ||:    ||:|::|
Db      1728 LWHTGGKGVDLVLNSLAEEKLQASVRCFGTHGRFLEIGKFDLSQNHPLGMAIFLKNVTFH 1787

Qy      1210 LVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKL 1269
                 | |      |    |   |  |  :  :     |   ||   |       |:|  ||||  |||  :|:||:
Db      1788 GVLLDAFFNESSADWREVWALVEAAIRDGVVRPLKCTVFHGAQVEDAFRYMAQGKHIGKV 1847

Qy      1270 VLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASA-----APAAR----AAALEAFLRTQ 1320
             |: :    |   :        || |            :| ||         || :       |    |    |     :
Db      1848 VVQVLAEEPAVL----KGAKP---------KLMSAISKTFCPAHKSYIIAGGLGGF-GLE 1893

Qy      1321 VSQVLRTPEIKVGAE--ALFTRLGM-DSLMAVELRNRIEASLKLKLSTTFLST 1370
             ::| |      |: |  :     | :| |:      |  ::|       |::::||: :|:
Db      1894 LAQWL----IQRGVQKLVLTSRSGIRTGYQAKQVRRWRRQGLQVQVSTSNISS 1942
```

RESULT 14
PKSK_BACSU
ID PKSK_BACSU STANDARD; PRT; 4447 AA.
AC P40803;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative polyketide synthase pksK (PKS).
GN PKSK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / PB1424;
RX MEDLINE=95219083; PubMed=7704258;
RA Albertini A.M., Caramori T., Scoffone F., Scotti C., Galizzi A.;
RT "Sequence around the 159 degree region of the Bacillus subtilis
RT genome: the pksX locus spans 33.6 kb.";
RL Microbiology 141:299-309(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC SECONDARY METABOLISM.
CC -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -!- SIMILARITY: Contains 5 acyl carrier domains.
CC --

CC --
DR EMBL; U11039; AAA85144.1; -.
DR EMBL; Z99112; CAB13590.1; -.
DR EMBL; Z99113; CAB13601.1; -.
DR PIR; A69679; A69679.
DR HSSP; P14687; 1AMU.
DR SubtiList; BG10930; pksK.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR006163; Pp_bind.
DR InterPro; IPR006162; Ppantne_attach.
DR Pfam; PF00501; AMP-binding; 1.
DR Pfam; PF00668; Condensation; 1.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR Pfam; PF00550; pp-binding; 5.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.

```
DR   PROSITE; PS00455; AMP_BINDING; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 5.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat; Ligase;
KW   Complete proteome.
FT   DOMAIN        1     68       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1063   1130       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2516   2589       ACYL CARRIER (ACP) 3.
FT   DOMAIN     2618   2687       ACYL CARRIER (ACP) 4.
FT   DOMAIN     3868   3937       ACYL CARRIER (ACP) 5.
FT   BINDING      31     31       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    1093   1093       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2552   2552       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2650   2650       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE   2915   2915       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   BINDING    3900   3900       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE   4147   4147       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
SQ   SEQUENCE   4447 AA;  496058 MW;  9C6DB4A6C0A9C057 CRC64;

  Query Match             14.1%;  Score 1018;  DB 1;  Length 4447;
  Best Local Similarity   26.3%;  Pred. No. 1.4e-45;
  Matches  391;  Conservative  203;  Mismatches  519;  Indels  376;  Gaps   52;

Qy          2 ADRPIERAAE-DPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWF-D 59
Db       2733 AERNKKQAADFEPVAIVGISGRFP-GAMDIDEFWKNLEEGKDSITEVPKDRWDWREHYGN 2791

Qy         60 PDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAP 119
Db       2792 PDTDV-NKTDIKWGGFIDGVAEFDPLFFGISPREADYVDPQQRLLMTYVWKALEDAGCSP 2850

Qy        120 SALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM-PSVGAGRISYALGLRGPC 178
Db       2851 QSLSGTGTGIFIGTGNTGYKDLFHR--ANLPIEGHAATGHMIPSVGPNRMSYFLNIHGPS 2908

Qy        179 VAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKA 238
Db       2909 EPVETACSSSLVAIHRAVTAMQNGDCEMAIAGGVNTILTEEAHISYSKAGMLSTDGRCKT 2968

Qy        239 FSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEI 298
Db       2969 FSADANGYVRGEGVGMVMLKKLEDAERDGNHIYGVIRGTAENHGGRANTLTSPNPKAQAD 3028

Qy        299 VLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVY---GLGRDVATPLL----- 350
Db       3029 LLVRAYRQADIDPSTVTYIEAHGTGTELGDPIEINGLKAAFKELSNMRGESQPDVPDHRC 3088

Qy        351 -IGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRART 409
Db       3089 GIGSVKSNIGHLELAAGISGLIKVLLQMKHKTLVKSLHCETLNPYLQLTDSPFYIVQEKQ 3148

Qy        410 PWP-----DWN-TPRRAGVSSFGMSGTNAHVVLEE-APAATCTPPAPERPAELLVLSART 462
Db       3149 EWKSVTDRDGNELPRRAGISSFGIGGVNAHIVIEEYMPKANSEHTATEQP-NVIVLSAKN 3207

Qy        463 ASALDAQAARLRDHL--ETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAA 520
Db       3208 KSRLIDRASQLLEVIRNKKYTDQDLHRIAYTLQVGREEMDERLACVAGTMQELEEKLQAF 3267

Qy        521 AQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQE 580
Db       3268 VDGKEETDEFFRGQSHRNKETQTIFTADEDMALAL-----DAWIRKRKYAKLA------- 3315

Qy        581 LDRPLREVMWAEPASVDAALLDQTAFTQPALFTF--------EYALAALWRSWGVEPELV 632
Db       3316 -------DLWVKGVSIQWNTL--YGETKPRLISLPSYPFAKDHYWVPAKEHSERDKKELV 3366

Qy        633 AGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPE-ADVAAAVA-- 689
Db       3367 --NAIEDRAACFLTKQWSLSP----------IGSAVPGTRTVAILCCQETADLAAEVSSY 3414

Qy        690 -PHAASVSIAAVNAPDQ-----------VVIAGAG----------QPVHAIAAAMAARGA 727
Db       3415 FPNHLLIDVSRIE-NDQSDIDWKEFDGLVDVIGCGWDDEGRLDWIEWVQRLVEFGHKEGL 3473

Qy        728 R----TKAL-------------------HVSHAFHSPLMAPMLEAFGRVAESVSYRRPSI 764
Db       3474 RLLCVTKGLESFQNTSVRMAGASRAGLYRMLQCEYSHLISRHMDA----EEVTDHRR--- 3526

Qy        765 VLVSNLSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVP 824
Db       3527 ------LAKLIADEFYSDSYDAE-----VCYRDGLR------YQAFLKAHPETGKATEQS 3569

Qy        825 ACMPDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQR 884
Db       3570 AVFP-------------KDH---VLLITGGTRGIGLLCA--------------------- 3592

Qy        885 ERYWIDTKADDAARGDRRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRR------- 937
Db       3593 -RHFAE------CYGVKKLVLTGREQL-----------------PPREEWARFKTSNTS 3627

Qy        938 --EKVEAAGDRPFRLEIDEPGVL----------DHLVLRVTERRAPGLGEVEIAVDAAGL 985
Db       3628 LAEKIQAV------RELEAKGVQVEMLSLTLSDDAQVEQTLQHIKRTLGPIGGVIHCAGL 3681

Qy        986 SFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFAT 1045
Db       3682 TDMDTLAFIRKTSDDIQRVLEPK-----------------VSGLT-----------TLYR 3713

Qy       1046 HVTTSAALVLPRPQALSAI--EAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGL 1103
Db       3714 HVCNEPLQFFVLFSSVSAIIPELSAGQADYAMANSYMDYFAEAHQKHAPII--------- 3764

Qy       1104 AAVQWA--QHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFVADVRAWTGGEGVDVV 1161
Db       3765 -SVQWPNWKETG------MGEVTNQAYRDS-GLLSITNSEGLRFLDQIVSKKFG---PVV 3813

Qy       1162 LNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERP 1221
Db       3814 LPAMA-------------------------NQTNWEPEL------------LMKRRK 3833

Qy       1222 ARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIR 1281
Db       3834 PHEGGLQE-------AALQSPPA------------RDIEEADEVSK------------- 3860

Qy       1282 IPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALFTRL 1341
Db       3861 ----------------CDGLLSET--------QSWLIDLFTEELRIDREDFEIDGLFQDY 3896

Qy       1342 GMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATALS 1390
Db       3897 GVDSIILAQVLQRINRKLEAALDPSILYEYPTIQRFADWLIGSYSERLS 3945

RESULT 15
FAS_CHICK
ID   FAS_CHICK      STANDARD;      PRT;  2511 AA.
AC   P12276;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39;
DE   EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN   FASN OR FAS.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RC   STRAIN=White leghorn; TISSUE=Liver;
RX   MEDLINE=95031085; PubMed=7944406;
RA   Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Amino-terminal blocking group and sequence of the animal fatty acid
```

```
RT   synthase.";
RL   Arch. Biochem. Biophys. 314:45-49(1994).
RN   [2]
RP   SEQUENCE OF 75-1775 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=89282777; PubMed=2734291;
RA   Holzer K.P., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of chicken liver fatty acid
RT   synthase cDNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:4387-4391(1989).
RN   [3]
RP   SEQUENCE OF 1568-2512 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89139426; PubMed=2917973;
RA   Chirala S.S., Kasturi R., Pazirandeh M., Stolow D.T., Huang W.-Y.,
RA   Wakil S.J.;
RT   "A novel cDNA extension procedure. Isolation of chicken fatty acid
RT   synthase cDNA clones.";
RL   J. Biol. Chem. 264:3750-3757(1989).
RN   [4]
RP   SEQUENCE OF 1752-2512 FROM N.A.
RX   MEDLINE=88320436; PubMed=2842766;
RA   Yuan Z., Liu W., Hammes G.G.;
RT   "Molecular cloning and sequencing of DNA complementary to chicken
RT   liver fatty acid synthase mRNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:6328-6331(1988).
RN   [5]
RP   SEQUENCE OF 2202-2512 FROM N.A.
RX   MEDLINE=89088152; PubMed=3207710;
RA   Kasturi R., Chirala S.S., Pazirandeh M., Wakil S.J.;
RT   "Characterization of a genomic and cDNA clone coding for the
RT   thioesterase domain and 3' noncoding region of the chicken liver
RT   fatty acid synthase gene.";
RL   Biochemistry 27:7778-7785(1988).
RN   [6]
RP   SEQUENCE OF 2121-2209.
RX   MEDLINE=89192401; PubMed=2648999;
RA   Huang W.-Y., Stoops J.K., Wakil S.J.;
RT   "Complete amino acid sequence of chicken liver acyl carrier protein
RT   derived from the fatty acid synthase.";
RL   Arch. Biochem. Biophys. 270:92-98(1989).
RN   [7]
RP   SEQUENCE OF 2209-2508.
RC   STRAIN=White leghorn;
RX   MEDLINE=89088151; PubMed=3207709;
RA   Yang C.-Y., Huang W.-Y., Chirala S.S., Wakil S.J.;
RT   "Complete amino acid sequence of the thioesterase domain of chicken
RT   liver fatty acid synthase.";
RL   Biochemistry 27:7773-7777(1988).
RN   [8]
RP   SEQUENCE OF 667-674 AND 1698-1709.
RX   MEDLINE=89323081; PubMed=2751995;
RA   Chang S.I., Hammes G.G.;
RT   "Amino acid sequences of pyridoxal 5'-phosphate binding sites and
RT   fluorescence resonance energy transfer in chicken liver fatty acid
RT   synthase.";
RL   Biochemistry 28:3781-3788(1989).
CC   -!- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF LONG-
CC       CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH. THIS
CC       MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN ACYL
CC       CARRIER PROTEIN.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC       long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC       acetyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC       malonyl-[acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC       carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC       [acyl-carrier protein].
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC       NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC       = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC   -!- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC       2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC   -!- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC       carrier protein] + oleate.
CC   -!- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=2;
CC         IsoId=P12276-1; Sequence=Displayed;
CC       Name=1;
CC         IsoId=P12276-2; Sequence=VSP_000149;
CC   -!- CAUTION: REF.3 AND REF.5 SEQUENCES DIFFER FROM THAT SHOWN FROM
CC       POSITION 2351 ONWARDS DUE TO A FRAMESHIFT.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J04485; AAB46389.1; -.
DR   EMBL; J03860; AAA48767.1; -.
DR   EMBL; J02839; AAA82106.1; ALT_SEQ.
DR   PIR; S57248; XYCHFA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00107; ADH_zinc_N; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Lyase; NADP; Acetylation;
KW   Alternative splicing; Pyridoxal phosphate.
FT   INIT_MET      0      0
FT   DOMAIN        1   ?411       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      427    815       ACYL AND MALONYL TRANSFERASES.
FT   DOMAIN     1638   1866       ENOYL REDUCTASE.
FT   DOMAIN     1867   2119       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2124   2180       ACYL CARRIER (ACP).
FT   DOMAIN     2209   2511       THIOESTERASE.
FT   MOD_RES       1      1       ACETYLATION.
FT   ACT_SITE    160    160       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    579    579       ACYL/MALONYL TRANSFERASES
FT                                (BY SIMILARITY).
FT   ACT_SITE    877    877       BETA-HYDROXYACYL DEHYDRATASE
FT                                (BY SIMILARITY).
FT   NP_BIND    1674   1691       NADP (ER).
FT   BINDING    1707   1707       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   NP_BIND    1888   1903       NADP (KR).
FT   BINDING    2157   2157       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2308   2308       THIOESTERASE (BY SIMILARITY).
FT   ACT_SITE   2481   2481       THIOESTERASE (BY SIMILARITY).
FT   VARSPLIC   2348   2348       T -> TQCFSFSLF (in isoform 1).
FT                                /FTId=VSP_000149.
FT   CONFLICT     77     78       QL -> PV (IN REF. 2).
FT   CONFLICT    116    116       L -> A (IN REF. 2).
FT   CONFLICT    675    675       R -> S (IN REF. 2).
FT   CONFLICT   1169   1169       K -> N (IN REF. 2).
FT   CONFLICT   1178   1178       A -> T (IN REF. 2).
FT   CONFLICT   1191   1191       R -> H (IN REF. 2).
FT   CONFLICT   1198   1198       P -> L (IN REF. 2).
FT   CONFLICT   1286   1287       DN -> ND (IN REF. 2).
```

```
FT   CONFLICT    1372   1372       K -> E (IN REF. 2).
FT   CONFLICT    1533   1533       C -> Y (IN REF. 2).
FT   CONFLICT    1577   1577       W -> R (IN REF. 2).
FT   CONFLICT    1685   1696       QAAIAIALSMGC -> ASSHCHRLEHGLA (IN REF. 2;
FT                                 AAA48767).
FT   CONFLICT    1732   1732       Q -> E (IN REF. 2).
FT   CONFLICT    1745   1745       S -> N (IN REF. 2).
SQ   SEQUENCE    2511 AA;  274648 MW;  622039DAC8315D3F CRC64;

  Query Match             13.8%;  Score 997.5;  DB 1;  Length 2511;
  Best Local Similarity   21.7%;  Pred. No. 8.6e-45;
  Matches  425;  Conservative  186;  Mismatches  564;  Indels  783;  Gaps   43;

Qy         12 DPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKTPVT 71
Db          1 EDVVIAGIAGKLPESE-NLQEFWENLLNGVDMVTE-DDRRWKPGIYGLP---------K 48

Qy         72 RASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGVFI 131
Db         49 RNGKLKDIKKFDASFFGVHPKQAHTMDPQLRLLLEVSYEAILDGGINPTALRGTDTGVWV 108

Qy        132 GIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSSLVA 191
Db        109 GASGSEALEALSQ--DPEELLGYSMTGCQRAMLANRISYFYDFTGPSLTIDTACSSSLMA 166

Qy        192 VHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGRGEG 251
Db        167 LENAYKAIRHGQCSAALVGGVNILLKPNTSVQFMKLGMLSPDGACKAFDVSGNGYCRSEA 226

Qy        252 CAVVVLKRLSGARADGDRILAVIRGSAINHDG-ASSGLTVPNGSSQEIVLKRALADAGCA 310
Db        227 VVVVLLTKKSMAK----RVYATIVNAGSNTDGFKEQGVTFPSGEMQQQLVGSLYRECGIK 282

Qy        311 ASSVGYVEAHGTGTTLGDPIEIQAL-NAVYGLGRDVATPLLIGSVKTNLGHPEYASGITG 369
Db        283 PGDVEYVEAHGTGTKVGDPQEVNGIVNVFCQCERE---PLLIGSTKSNMGHPEPASGLAA 339

Qy        370 LLKVVLSLQHGQIPAHLHAQALNPRISW---GDLRLTVTRARTPWPDWNTPRRAG---VS 423
Db        340 LAKVILSLEHGLWAPNLHFNDPNPDIPALHDGSLKVVCK---------PTPVKGGLVSIN 390

Qy        424 SFGMSGTNAHVVLEE-----APAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLE 478
Db        391 SFGFGGSNAHVILRPNEKKCQPQETCNLP------RLVQVCGRTQEAVEILIEESRKHGG 444

Qy        479 TYP-SQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIAD-- 535
Db        445 CSPFLSLLSDIS---AVPVSSMPYR-------------------GYTLVGTESDITEIQ 481

Qy        536 ---SSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAE 592
Db        482 QVQASGRPLWYICSGMGTQWKGMGLSL------------MKLDLFRQSILRS-DEALKST 528

Qy        593 PASVDAALL-------DQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACV 645
Db        529 GLKVSDLLLNADENTFDDTVHAFVGLAAIQIAQIDVLKAAGLQPDGILGHSVGELACGYA 588

Qy        646 AGVFSLEDAVFLVAARGRLMQ--ALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAP 703
Db        589 DNSLSHEEAVLAAYWRGRCVKEAKLPPGG-MAAVGLTWEECKQRCPP---NVVPACHNSE 644

Qy        704 DQVVIAGAGQPVHAIAAAMAARGARTKAL-HVSHAFHSPLMAPMLEAF----------- 750
Db        645 DTVTVSGPLDSVSEFVTKLKKDGVFAKEVRRAGVAFHSYYMASIAPALLSALKKVIPHPK 704

Qy        751 GRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTF 810
Db        705 PRSARWISTSIPESQWQSDLARN------SSAEYHVNNLVNPVLFHEGLK--HIPENAVV 756

Qy        811 VEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPS 870
Db        757 VEIAPHALLQAILRRTLKPTCTILPLMKKDHKNNLEFFLTQTGKIHLTGINVLGNNLFPP 816

Qy        871 GGRRVPL------PTYPWQRERYWIDTKADDAARGDRRAPGA----------------G 907
Db        817 VEYPVPVGTPLISPYIKWDHSQDWDVPKAEDFPSGSKGSASASVYNIDVSPDSPDHYLVG 876

Qy        908 H----------------------------------------------------------- 908
Db        877 HCIDGRVLYPATGYLVLAWRTLARSLGMVMEQTAVMFEEVTIHQATILPKKGSTQLEVRI 936

Qy        909 ------------------------------------------------------------ 908
Db        937 MPASHSFEVSGNGNLAVSGKISLLENDALKNFHNQLADFQSQANVTAKSGLLMEDVYQEL 996

Qy        909 ------------------------------------------------------------ 908
Db        997 HLRGYNYGPTFQGVLECNSEGSAGKILWNGNWVTFLDTLLHLIVLAETGRSLRLPTRIRS 1056

Qy        909 ----------------------------DEVEEGG------------------------- 915
Db       1057 VYIDPVLHQEQVYQYQDNVEAFDVVVDRCLDSLKAGGVQINGLHASVAPRRQQERISPTL 1116

Qy        916 ---------------------------------------------AVRGGDRRSARLDHP 930
Db       1117 EKFSFVPYIESDCLSSSTQLHAYLEHCKGLIQKLQAKMALHGVKLVIHGLETKGAAAGSP 1176

Qy        931 PPESGRRE------KVEAAGDRPFRLE----------IDEP---GVLDHLVLRV------ 965
Db       1177 PAQKGLQHILTEICRLELNGNPHSELEQIVTQEKMHLQDDPLLNGLLDSSELKTCLDVAK 1236

Qy        966 -----------------------------------------------------TERRAPG- 972
Db       1237 ENTTSHRMKIVEALAGSGRLFSRVQSILNTQPLLQLDYIATDCTPETLSDNETELHDAGI 1296

Qy        973 ------------------------------------------------------------ 972
Db       1297 SFSQWDPSSLPSGNLTNADLAVCNCSTSVLGNTAEIISNLAAAVKEGGFVLLHTLLKEET 1356

Qy        973 LGEV-------------------------------------------------------- 976
Db       1357 LGEIVSFLTSPDLQQKHSFLSQAQWEELFSKASLNLVAMKRSFFGSVIFLCRRQSPAKAP 1416

Qy        977 -----------------EIAVDA------------------------------------- 982
Db       1417 ILLPVDDTHYKWVDSLKEILADSSEQPLWLTATNCGNSGILGMVNCLRLEAEGHRIRCVF 1476

Qy        983 ------------------------------------------------------------ 982
Db       1477 VSNLSPSSTVPATSLSSLEMQKIIERDLVMNVYRDGKWGSFRHLPLQQAQPQELTECAYV 1536

Qy        983 ------------------------------------AGLSFNDVQLALG-MVPDDLPGK-P 1005
Db       1537 NVLTRGDLSSLRWIVSPLRHFQTTNPNVQLCKVYYASLNFWDIMLATGKLSPDAIPGNWT 1596

Qy       1006 NPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIE 1065
Db       1597 LQQCMLGMEFSGRDLA-GRRVMGL--------LPAKGLATVVDCDKRFLWEVPENWTLEE 1647

Qy       1066 AAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEK 1125
Db       1648 AASVPVVYATAYYALVVRGGMKKGESVLIHSGSGGVGQAAIAIALSMGCRVFATVGSAEK 1707

Qy       1126 RAYLES----LGVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHG 1181
Db       1708 REYLQARFPQLDANSFASSRNTTFQQHILRVTNGKGVSLVLNSLAEEKLQASLRCLAQHG 1767

Qy       1182 RFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFT 1241
Db       1768 RFLEIGKFDLSNNSQLGMALFLKNVAFHGILLDSIFEEGNQEWEVVSELLTKGIKDGVVK 1827

Qy       1242 PPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQ 1279
Db       1828 PLRTTVFGKEEVEAAFRFMAQGKHIGKVMIKIQEEEKQ 1865
```

Search completed: October 2, 2003, 17:28:40
Job time : 54.2576 secs

```
OM protein - protein search, using sw model

Run on:          October  2, 2003, 17:27:29 ; Search time 26.2576 Seconds
                                           (without alignments)
                                           2544.975 Million cell updates/sec

Title:           US-10-014-717-2
Perfect score:   7210
Sequence:        1 VADRPIERAAEDPIAIVGAS..........GVQNDFVSSGADQDWEIIAL 1421

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:      127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_41:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2107.5	29.2	3567	1	ERY2_SACER	Q03132 saccharopol
2	1925.5	26.7	3491	1	ERY1_SACER	Q03131 saccharopol
3	1889	26.2	3519	1	OL56_STRAT	Q07017 streptomyce
4	1875	26.0	3172	1	ERY3_SACER	Q03133 saccharopol
5	1852.5	25.7	2110	1	MCAS_MYCBO	Q02251 mycobacteri
6	1592.5	22.1	1538	1	PPSB_MYCTU	Q10978 mycobacteri
7	1583	22.0	1876	1	PPSA_MYCTU	Q10977 mycobacteri
8	1180	16.4	1986	1	WA_EMENI	Q03149 emericella
9	1177	16.3	1774	1	MSAS_PENPA	P22367 penicillium
10	1173.5	16.3	2109	1	PKS1_ASPPA	Q12053 aspergillus
11	1100	15.3	2505	1	FAS_RAT	P12785 rattus norv
12	1083	15.0	2181	1	STCA_EMENI	Q12397 emericella
13	1039	14.4	2504	1	FAS_HUMAN	P49327 homo sapien
14	1018	14.1	4447	1	PKSK_BACSU	P40803 bacillus su
15	997.5	13.8	2511	1	FAS_CHICK	P12276 gallus gall
16	940	13.0	4427	1	PKSL_BACSU	Q05470 bacillus su
17	882.5	12.2	4273	1	PKSM_BACSU	P40872 bacillus su
18	612.5	8.5	352	1	FAS_ANSAN	P36189 anser anser
19	392.5	5.4	317	1	FABD_BACSU	P71019 bacillus su
20	363	5.0	401	1	NOE4_RHIME	P06231 rhizobium m
21	354.5	4.9	402	1	NODE_RHIME	P06230 rhizobium m
22	344	4.8	308	1	FABD_SALTY	O85140 salmonella
23	336	4.7	403	1	NODE_RHILV	P04683 rhizobium l
24	332	4.6	308	1	FABD_ECOLI	P25715 escherichia
25	329.5	4.6	467	1	KASA_STRCO	Q02059 streptomyce
26	328	4.5	312	1	FABD_HAEIN	P43712 haemophilus
27	325.5	4.5	402	1	NODE_RHIS3	P72331 rhizobium s
28	324.5	4.5	401	1	NODE_RHILT	P04684 rhizobium l
29	321.5	4.5	426	1	KAS1_STRGA	P16538 streptomyce
30	316.5	4.4	300	1	VAT1_HUMAN	Q99536 homo sapien
31	315	4.4	412	1	FABF_ECOLI	P39435 escherichia
32	308	4.3	414	1	FABF_VIBHA	P55338 vibrio harv
33	304.5	4.2	293	1	FABD_SYNY3	P73242 synechocyst
34	303.5	4.2	413	1	FABF_VIBCH	Q9kqh9 vibrio chol
35	303	4.2	838	1	FAS_MOUSE	P19096 mus musculu
36	302	4.2	420	1	KAS1_STRCM	P41175 streptomyce
37	298.5	4.1	327	1	QOR_ECOLI	P28304 escherichia
38	298	4.1	421	1	KAS1_STRVN	P16540 streptomyce
39	295.5	4.1	327	1	QOR_SALTY	P40783 salmonella
40	293	4.1	421	1	FABF_RHIME	P56902 rhizobium m
41	289.5	4.0	422	1	KAS1_STRHA	Q05356 streptomyce
42	285.5	4.0	416	1	FABF_SYNY3	P73283 synechocyst
43	284.5	3.9	423	1	KAS1_STRCO	P23155 streptomyce
44	282	3.9	379	1	VAT1_TORCA	P19333 torpedo cal
45	278	3.9	325	1	QOR_PSEAE	P43903 pseudomonas

ALIGNMENTS

```
RESULT 1
ERY2_SACER
ID   ERY2_SACER     STANDARD;      PRT;  3567 AA.
AC   Q03132; Q54096;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE   deoxyerythronolide B synthase II) (DEBS 2).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: Contains 2 acyl carrier domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M63677; AAA26494.1; -.
DR   EMBL; X62569; CAA44448.1; -.
DR   PIR; S23070; S23070.
DR   InterPro; IPR001227; Ac_transferase..
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00107; ADH_zinc_N; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 3.
FT   DOMAIN     1485   3567       MODULE 4.
FT   DOMAIN       27    488       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      559    884       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1130   1301       BETA-KETOACYL REDUCTASE 1 (POSSIBLY
FT                                NON-FUNCTIONAL).
FT   DOMAIN     1397   1467       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1485   1943       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2013   2336       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2383   3066       DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT   DOMAIN     3139   3322       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3415   3485       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    202    202       THIOESTER BOND.
FT   ACT_SITE    651    651       ACYL-ESTER INTERMEDIATE.
FT   BINDING    1430   1430       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2105   2105       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    2961   2978       NADP (ER).
FT   NP_BIND    3142   3157       NADP (KR).
FT   BINDING    3448   3448       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    438    438       R -> A (IN REF. 2).
FT   CONFLICT    480    480       T -> S (IN REF. 2).
FT   CONFLICT   1241   1241       L -> F (IN REF. 2).
FT   CONFLICT   2664   2664       G -> V (IN REF. 2).
SQ   SEQUENCE   3567 AA;  374413 MW;  EE6284F4738AA0C0 CRC64;

  Query Match             29.2%;  Score 2107.5;  DB 1;  Length 3567;
  Best Local Similarity   32.5%;  Pred. No. 4.5e-103;
  Matches  610;  Conservative  179;  Mismatches  530;  Indels  559;  Gaps   45;

Qy          2 ADRPIERAAE-DPIAIVGASCRLPGGVIDLSGFWTLL-EGSRDTVGRVPAER-WDAAAWF 58
Db       1478 AVRVVGAADESEPIAIVGIGCRFPGGIGSPEQLWRVLAEGANLTTG-FPADRGWDIGRLY 1536

Qy         59 DPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIA 118
Db       1537 HPDPDNPGTSYVDKGGFLTDAADFDPGFFGITPREALAMDPQQRLMLETAWEAVERAGID 1596

Qy        119 PSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPC 178
Db       1597 PDALRGTDTGVFVGMNGQSYMQLL--AGEAERVDGYQGLGNSASVLSGRIAYTFGWEGPA 1654

Qy        179 VAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKA 238
Db       1655 LTVDTACSSSLVGIHLAMQALRRGECSLALAGGVTVMSDPYTFVDFSTQRGLASDGRCKA 1714

Qy        239 FSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEI 298
Db       1715 FSARADGFALSEGVAALVLEPLSRARANGHQVLAVLRGSAVNQDGASNGLAAPNGPSQER 1774

Qy        299 VLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNL 358
Db       1775 VIRQALAASGVPAADVDVVEAHGTGTELGDPIEAGALIATYGQDRD--RPLRLGSVKTNI 1832

Qy        359 GHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPR 418
Db       1833 GHTQAAAGAAGVIKVVLAMRHGMLPRSLHADELSPHIDWESGAVEVLREEVPWPAGERPR 1892

Qy        419 RAGVSSFGMSGTNAHVVLEEAP-----AATCTPPAPERPAELLVLSARTASALDAQAARL 473
Db       1893 RAGVSSFGVSGTNAHVIVEEAPAEQEAARTERGPLP------FVLSGRSEAVVAAQARAL 1946

Qy        474 RDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSI 533
Db       1947 AEHLRDTPELGLTDAAWTLATGRARFDVRAAVLGDDRAGVCAELDALAEGRPSADAVAPV 2006

Qy        534 ADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEP 593
Db       2007 TSAPR-KPVLVFPGQGAQWVGMARDLLESSEVFAESMSRCAEALSPHTDWKLLDVVRGDG 2065

Qy        594 ASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLED 653
Db       2066 GPDPHERVD---VLQPVLFSIMVSLAELWRAHGVTPAAVVGHSQGEIAAAHVAGALSLEA 2122

Qy        654 AVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQ 713
Db       2123 AAKVVALRSQVLRELDDQGGMVSVGASRDELETVLARWDGRVAVAAVNGPGTSVVAGPTA 2182

Qy        714 PVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGK 773
Db       2183 ELDEFFAEAEAREMKPRRIAVRYASHSPEVARIEDRLAAELGTITAVRGSVPLHSTVTGE 2242

Qy        774 ACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLL----------GLV 823
Db       2243 VIDTSAMDASYWYRNLRRPVLFEQAVRGLVEQGFDTFVEVSPHPVLLMAVEETAEHAGAE 2302

Qy        824 PACMPDARPALLASSRAGRDEPATVLEAL--GGLWAVGGLVSWAGLFP--SGGRRVPLPT 879
Db       2303 VTCVP--------TLRREQSGPHEFLRNLLRAHVHGVG-----ADLRPAVAGGRPAELPT 2349

Qy        880 YPWQRERYW--IDTKADDAARGDRRA---------------------------------- 903
Db       2350 YPFEHQRFWPRPHRPADVSALGVRGAEHPLLLAAVDVPGHGGAVFTGRLSTDEQPWLAEH 2409

Qy        904 --------PG---------AGHD------------------------------------- 909
Db       2410 VVGGRTLVPGSVLVDLALAAGEDVGLPVLEELVLQRPLVLAGAGALLRMSVGAPDESGRR 2469

Qy        910 ------------------------------------------------------------ 909
Db       2470 TIDVHAAEDVADLADAQWSQHATGTLAQGVAAGPRDTEQWPPEDAVRIPLDDHYDGLAEQ 2529

Qy        910 ----------------------EV-----EEG-----------------GAV------- 917
Db       2530 GYEYGPSFQALRAAWRKDDSVYAEVSIAADEEGYAFHPVLLDAVAQTLSLGALGEPGGGK 2589

Qy        918 ------------------------RGGDRRSARL----------------------DHP 930
Db       2590 LPFAWNTVTLHASGATSVRVVATPAGADAMALRVTDPAGHLVATVDSLVVRSTGEKWEQP 2649

Qy        931 PPESGRRE-------------------------------------------KVEAAGDRP- 947
Db       2650 EPRGGEGELHALDWGRLAEPGSTGRVVAADASDLDAVLRSGEPEPDAVLVRYEPEGDDPR 2709

Qy        948 ------------------------------------------------------------ 947
Db       2710 AAARHGVLWAAALVRRWLEQEELPGATLVIATSGAVTVSDDDSVPEPGAAAMWGVIRCAQ 2769

Qy        948 -------FRLEID-EPGVL---------------DHLVLRV-------------TERRAPG 972
Db       2770 AESPDRFVLLDTDAEPGMLPAVPDNPQLALRGDDVFVPRLSPLAPSALTLPAGTQRLVPG 2829

Qy        973 --------------------LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012
Db       2830 DGAIDSVAFEPAPDVEQPLRAGEVRVDVRATGVNFRDVLLALGMYPQKAD--------MG 2881
```

```
Qy       1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072
              | ||  : |||  |:      |  |: |  ||||      |     |:     |    :|||:|:|
Db       2882 TEAAGVVTAVGPDVDAFAPGDRVLGLFQGAFAPIAVTDHRLLARVPDGWSDADAAAVPIA 2941

Qy       1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132
              | || |||   :|  |:  |:  ||||| ||||:|||   |:   |||| |||| | |    | :|
Db       2942 YTTAHYALHDLAGLRAGQSVLIHAAAGGVGMAAVALARRAGAEVLATAG-PAKHGTLRAL 3000

Qy       1133 GV--RYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRD 1190
              |:    ::: ||    |     |    ||| ||||||||:|||:|:|  :||    |  |||:|| |
Db       3001 GLDDEHIASSRETGFARKFRERTGGRGVDVVLNSLTGELLDESADLLAEDGVFVEMGKTD 3060

Qy       1191 CYADNQLGLRPFLRNLSFSLVDLRGMMLERPA----------RVRALLEELLGLIAAGVF 1240
                               ||:      | ||     | |           |:  :|  |::||: ||
Db       3061 ------------LRDAG----DFRG----RYAPFDLGEAGDDRLGEILREVVGLLGAGEL 3100

Qy       1241 TPPPIATLPIARVADAFRSMAQAQHLGKLVLTLG---DPEVQIRIPTHAGAGPSTGDRDL 1297
                 |::    :       | :  |:: :|:||||||      ||:  :  |                ||
Db       3101 DRLPVSAWELGSAPAALQHMSRGRHVGKLVLTQPAPVDPDGTVLI---------TGGTGT 3151

Qy       1298 LDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEA 1357
              | ||        ||      |   :|  :                     | :| | |:   : |||  ||
Db       3152 LGRL-----LARHLVTEHGVRHLL---------------LVSRRGADAPGSDELRAEIE- 3190

Qy       1358 SLKLKLSTTFLSTSPNIALLAQNLLDALATAL---------------SLERVAAEN-LR 1400
                 |            |:   || |  ||| |   |                    |::  |  |   ||
Db       3191 DLGASAEIAACDTADRDALSA--LLDGLPRPLTGVVHAAGVLADGLVTSIDEPAVEQVLR 3248

Qy       1401 AGVQNDFVSSGADQDWEI 1418
              | |          |  | :
Db       3249 AKV---------DAAWNL 3257
```

```
RESULT 2
ERY1_SACER
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
OX   NCBI_TaxID=1836;.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: Contains 3 acyl carrier domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   PIR; T43231; T43231.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1972       MODULE 1.
FT   DOMAIN     1979   3491       MODULE 2.
FT   DOMAIN        1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN      503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1888   1958       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN     3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3334   3404       ACYL CARRIER (ACP) 3.
FT   ACT_SITE    145    145       ACYL-ESTER INTERMEDIATE.
FT   BINDING     447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   1128   1128       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    1614   1660       NADP.
FT   BINDING    1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2598   2598       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    3058   3104       NADP.
FT   BINDING    3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match             26.7%;  Score 1925.5;  DB 1;  Length 3491;
  Best Local Similarity   35.6%;  Pred. No. 1.7e-93;
  Matches  547;  Conservative  184;  Mismatches  601;  Indels  203;  Gaps   39;

Qy          4 RPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDP 62
              |   |  |    :|:|:|   :||||||        || ||      ||  |    :|  |:  |   : |  |||
Db        497 RTNEAAPGEPVAVVAMACRLPGGVSTPEEFWELLSEGRDAVAGLPTDRGWDLDSLFHPDP 556

Qy         63 DAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSAL 122
                        |       ||::     ||  :|||:||||||  :||    ||:||:  || ||  |  |  |::|
Db        557 TRSGTAHQRGGGFLTEATAFDPAFFGMSPREALAVDPQQRLMLELSWEVLERAGIPPTSL 616
```

```
Qy   123 VGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVD 182
Db   617 QASPTGVFVGLIPQEYGPRL--AEGGEGVEGYLMTGTTTSVASGRIAYTLGLEGPAISVD 674

Qy   183 TAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAE 242
Db   675 TACSSSLVAVHLACQSLRRGESSLAMAGGVTVMPTPGMLVDFSRMNSLAPDGRCKAFSAG 734

Qy   243 ADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKR 302
Db   735 ANGFGMAEGAGMLLLERLSDARRNGHPVLAVLRGTAVNSDGASNGLSAPNGRAQVRVIQQ 794

Qy   303 ALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPE 362
Db   795 ALAESGLGPADIDAVEAHGTGTRLGDPIEARALFEAY--GRDREQPLHLGSVKSNLGHTQ 852

Qy   363 YASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGV 422
Db   853 AAAGVAGVIKMVLAMRAGTLPRTLHASERSKEIDWSSGAISLLDEPEPWPAGARPRRAGV 912

Qy   423 SSFGMSGTNAHVVLEEAP---------AATCTPPAPERPAELLVLSARTASALDAQAARL 473
Db   913 SSFGISGTNAHAIIEEAPQVVEGERVEAGDVVAP--------WVLSASSAEGLRAQAARL 964

Qy   474 RDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSI 533
Db   965 AAHLREHPGQDPRDIAYSLATGRAALPHRAAFAPVDESAALRVLDGLATGNADGAAVGT- 1023

Qy   534 ADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEP 593
Db  1024 -SRAQQRAVFVFPGQGWQWAGMAVDLLDTSPVFAAALRECADALEPHLDFEVIPFLRAEA 1082

Qy   594 A--SVDAAL-LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFS 650
Db  1083 ARREQDAALSTERVDVVQPVMFAVMVSLASMWRAHGVEPAAVIGHSQGEIAAACVAGALS 1142

Qy   651 LEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAG 710
Db  1143 LDDAARVVALRSRVIATMPGNKGMASIAAPAGEVRARIGDR---VEIAAVNGPRSVVVAG 1199

Qy   711 AGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIV-LVSN 769
Db  1200 DSDELDRLVASCTTECIRAKRLAVDYASHSSHVETIRDALHAELGEDFHPLPGFVPFFST 1259

Qy   770 LSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPD 829
Db  1260 VTGRWTQPDELDAGYWYRNLRRTVRFADAVRALAEQGYRTFLEVSAHPILTAAIEEIGDG 1319

Qy   830 ARPALLA--SSRAGRDEPATVLEALGGLWAVGGLVSWAGL-FPSGGRRVPLPTYPWQRER 886
Db  1320 SGADLSAIHSLRRGDGSLADFGEALSRAFAAGVAVDWESVHLGTGARRVPLPTYPFQRER 1379

Qy   887 YWIDTKADDAARGDRRAPGAGHDEVEEGGAVR-------GGDRRSARLD----------- 928
Db  1380 VWLEPK--PVARRS--------TEVDEVSALRYRIEWRPTGAGEPARLDGTWLVAKYAGT 1429

Qy   929 HPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPGLGEVE-----IAVDAA 983
Db  1430 ADETSTAAREALESAGARVRELVVDARCGRDELA-----ERLRSVGEVAGVLSLLAVDEA 1484

Qy   984 GLSFNDVQLALGMVPDDLP----------GKPNPPLLLGGECAGRIVAVGEGVNGLV--V 1031
Db  1485 --EPEEAPLALASLADTLSLVQAMVSAELGCPLWTVTESAVATGPFERVRNAAHGALWGV 1542

Qy  1032 GQPVIALSAGAFATHVTTSAALVLPRPQALSAIE---------AAAMPVAYLTAWYALDR 1082
Db  1543 GRVIALENPAVWGGLVDVPAGSVAELARHLAAVVSGGAGEDQLALRADGVYGRRWVRAAA 1602

Qy  1083 IA---RLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHA---------TAGTPEKRAYLE 1130
Db  1603 PATDDEWKPTGTVLVTGGTGGVGGQIARWLARRGAP-HLLLVSRSGPDADGAGELVAELE 1661

Qy  1131 SLGVRY------VSDSRSDRFVADVRAWTGGEGVDVVL---------------NSLSGEL 1169
Db  1662 ALGARTTVAACDVTDRES------VRELLGGIGDDVPLSAVFHAAATLDDGTVDTLTGER 1715

Qy  1170 IDKSFNLLRSHGRFVELGKRDCY-ADNQLGLRPFLRNLSFSL---------------VDLR 1214
Db  1716 IERA-------SRAKVLGARNLHELTRELDLTAFVLFSSFASAFGAPGLGGYAPGNAYLD 1768

Qy  1215 GMMLER-----PARVRA---------LLEELLGLIA-----AGVFTPPP------------ 1244
Db  1769 GLAQQRRSDGLPATAVAWGTWAGSGMAE---GAVADRFRRHGVIEMPPETACRALQNALD 1825

Qy  1245 ---IATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRL 1301
Db  1826 RAEVCPIVIDVRWDRFLLAYTAQRPTRLFDEIDDAR---RAAPQAPAEPRVG------AL 1876

Qy  1302 ASAAPAARAAALEAFLRTQVSQVL-RTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLK 1360
Db  1877 ASLPAPEREEALFELVRSHAAAVLGHASAERVPADQAFAELGVDSLSALELRNRLGAATG 1936

Qy  1361 LKLSTTFLSTSPNIALLAQNLLDALATALSLERVA 1395
Db  1937 VRLPTTTVFDHPDVRTLAAHLAAELGGATGAEQAA 1971
```

RESULT 3

OL56_STRAT

```
ID   OL56_STRAT     STANDARD;      PRT;  3519 AA.
AC   Q07017;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Oleandomycin polyketide synthase, modules 5 and 6.
GN   ORFB.
OS   Streptomyces antibioticus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150470; PubMed=8107683;
RA   Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT   "Characterisation of a Streptomyces antibioticus gene encoding a type
RT   I polyketide synthase which has an unusual coding sequence.";
RL   Mol. Gen. Genet. 242:358-362(1994).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC       LACTONE RING.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- SIMILARITY: Contains 2 acyl carrier domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L09654; AAA19695.1; -.
DR   PIR; S43048; S43048.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
```

```
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN        1      ?       MODULE 5.
FT   DOMAIN        ?   3519       MODULE 6.
FT   DOMAIN       32    501       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      569    890       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1200   1382       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1487   1561       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1686   2156       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2220   2541       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2856   3038       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3141   3215       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3270   3519       THIOESTERASE.
FT   ACT_SITE    210    210       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE    660    660       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    1203   1249       NADP.
FT   BINDING    1524   1524       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1859   1859       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2311   2311       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    2859   2905       NADP.
FT   BINDING    3178   3178       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3519 AA;  368561 MW;  41AE78AAAEB61F86 CRC64;

  Query Match             26.2%;  Score 1889;  DB 1;  Length 3519;
  Best Local Similarity   34.0%;  Pred. No. 1.4e-91;
  Matches  541;  Conservative  191;  Mismatches  595;  Indels  264;  Gaps   44;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGK 67
              |: : |||  |||  ||:      |  |    || |  :| :| ||   | : |||: ||
Db       1684 ASREAIAITAMSCRFGGGIDSPEDLWRFLAEGRDAVAGLPEDRGWDLDALYHPDPENPGT 1743

Qy         68 TPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTET 127
              ||   :|  |  | ||| ||||||||| |||  |||||  ||  | | |  : || |
Db       1744 TYVREGAFRYDAAQFDAGFFGISPREALAMDPQQRLLLETSWELFERADIDPYTVRGTAT 1803

Qy        128 GVFIGIGPSEY---EAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184
              |:||| |   |        |:: |    :      ||  :| :|||| || || | ||||
Db       1804 GIFIGAGHQGYGPDPKRAPESVAGYLL-----TGTASAVLSGRISYTFGLEGPAVTVDTA 1858

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244
              |||||:||| |:|| |||| |:|||::| :| |   |  : :||||||||:| ||
Db       1859 CSSSLVALHLAVQALRRGECSLAIAGGVAVMSTPDAFVEFSRQQGMARDGRCKAFAAAAD 1918

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304
              | | |||  ::::|:||| ||   | :|||:||||:| ||||:||  ||| ||: |:: ||
Db       1919 GMGWGEGVSLLLLERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAAL 1978

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364
              |||| | : |  |||||||| |||||| ||| || | ||      | :||||:|:|| : |
Db       1979 ADAGLAPADVDVVEAHGTGTRLGDPIEAQALLATYGQGRAGGRPVWLGSVKSNIGHTQAA 2038

Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPW-PDWNTPRRAGVS 423
              :|: |::|:||:|  | :|  ||   :| : |   : :      || |:     ||||:|
Db       2039 AGVAGVMKMVLALGRGVVPKTLHVDEPSPHVDWSAGAVELLTEERPWEPEAERLRRAGIS 2098

Qy        424 SFGMSGTNAHVVLEEAPAATCTPPAPERPAELL---------VLSARTASALDAQAARLR 474
              :||:|||||||::||||     ||    ::              |:|  |   || |||||
Db       2099 AFGVSGTNAHVIVEEAPA----EPEPEPGTRVVAAGDLVVPWVVSGRDVGALREQAARLA 2154

Qy        475 DHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIA 534
              |: :  :  : || :||  |||  |||   :    :  | |  :  : :         :|
Db       2155 AHVSSTGAGVV-DVGWSLVATRSVFEHRAVMVGTDLDSMAGSLAGFAAGGVVPGVVSGVA 2213

Qy        535 DSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPA 594
               :   :: |:| |||:| :||  ||  |   ||  |  | : :      ||:
Db       2214 PAEGRRVVFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLQGR-- 2271

Qy        595 SVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDA 654
               || :| :    ||||:   :||  || :||||  | ||  ||: |||||  || |
Db       2272 --DATVLGRVDVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADG 2329

Qy        655 VFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQP 714
              :|  |  | :  :  || |||:  |   |    :   :   ||:||||  |  |::|  |
Db       2330 ARVVVLRSRAIARIAGGGGMVSVSLPAGRVRTMLDTYGGRVSVAAVNGPSSTVVSGDVQA 2389

Qy        715 VHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKA 774
              : : |        | | : : | :| ||  | |  : :            :: :  |:  |  ::
Db       2390 LDELLAGCEREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQDSSVPFFSTVTADW 2449

Qy        775 CTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMP--DARP 832
                   |||  :  || ||| : |: |  | | | | |||  |    |:  :   :  |
Db       2450 LDTTALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQNA 2509

Qy        833 ALLASSRAGRDEPA--TVLEALGGLWAVGGLVSWA----GLFPSGGRRVPLPTYPWQRER 886
              |:|  |  |  |||    :| :|  :  |   |  |   |: |   |  |||||:||:|
Db       2510 AVLGSLR--RDEGGLDRLLTSLAEAFVQGVPVDWTHAFEGVTP---RTVDLPTYPFQRQR 2564

Qy        887 YWIDTKADDAARG-DRRAPGAGHDEVEEGGAVR-----GGDRRSARLDHPPPESGRREKV 940
              :|:|     :|  |  |  |   | ||  :|      | |  :   |   |  |||  :|
Db       2565 FWLDGSPASSANGVDGEADAMIWDAVEREDSVAVAEELGIDAEALHTVLPALSSWRRRRV 2624

Qy        941 E---------AAGDRPPFRLEIDE---------------PGVLDHLVLRVTERRAPGLGE 975
              |          :||  :||                    ||:  :|  ||  |      ||
Db       2625 EHRRLQDWRYRVEWKPFPAALDEVLGGGWLFVVPRGLADDGVVARVVAAVTARG----GE 2680

Qy        976 VEI---------------AVDAAGLSFNDVQLALGMVP----DDLPGKPNPPLLLGGECA 1016
              | :            ||   |:|       |:|      ||      : |::  |  |
Db       2681 VSVVELDPTRPDRRAYAEAVAGRGVS--------GVVSFLSWDDRRHSEH-PVVPAGLAA 2731

Qy       1017 GRIVA--------VGEGVN-GLVVGQPVIA--LSAGAFATHVTTSA-----ALVLPRPQ- 1059
               ::|        ||||   ||     |:|     |||    |        |  | |:
Db       2732 SLVLAQALVDLGRVGEGPRLWLVTRDAVVAGPSDAGAVIDPVQAQVWGFGRVLGLEHPEL 2791

Qy       1060 ------------------------ALSAIEAAA-----MPVAYLTAWYALDRIARL---- 1086
                                      |: : |:|      :    |    |: |: |
Db       2792 WGGLIDLPVEAPEPGSTCDHTYADLLATVVASAGFEDQVAVRGSGVW--VRRLVRAVVDG 2849

Qy       1087 -----QPGERVLIHAATGGVGLAAVQWAQHVGAEVH---------ATAGTPEKRAYLESL 1132
                   :|    ||:    ||:|    :|     ||: |                  :   |  :  ||
Db       2850 GGGGWRPRGTVLVTGGLGGLGAHTARWLVGGGAD-HVVLVSRRGGSAPGAGDLVRELEGL 2908

Qy       1133 GVRYVSDSRSDRFVAD---VRAWTG--GEGVDVVLNSLS-------------------- 1166
               |   ||     |   |||    :||     ||  |   | ::
Db       2909 GGARVSVRACD--VADRVALRALLSDLGEPVTAVFHAAGVPQSTPLAEISVQEAADVMAA 2966

Qy       1167 --------GELID----KSFNLLRSHGRFVELGKRDCYA-------------------- 1193
                      |||:|     ::| |  |:      |  :  ||
Db       2967 KVAGAVNLGELVDPCGLEAFVLFSSNAGVWGSGGQAVYAAANAFLDALAVRRRGVGLPAT 3026

Qy       1194 ---------DNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPP 1244
                       :     :  ||      ||:      |  |  |:  :  : :|    |
Db       3027 SVAWGMWAGEGMASVGGAARELS-----RRGVRAMDPERAVAVMADAVGRGEAFV----A 3077

Qy       1245 IATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRI------PTHAGAGPSTGDRDLL 1298
               :|:   |    |  |      :   |        |||:   :          ||    |   ||
Db       3078 VADVDWERFVTGFASARPRPLISDL------PEVRTALRNQEQEQLHAPV-PEDRSAQLL 3130

Qy       1299 DRLASAAPAARAAALEAFLRTQVSQVLRTPEIK-VGAEALFTRLGMDSLMAVELRNRIEA 1357
               ||:  :||  | | |     :||: :  ||         : | ||   |   ||  |||  ||:|||||:  |
Db       3131 RRLSMLSPAGREAELVKLVRTEAAAVLGHGSAQDVPAERAFKELGFDSLTAVQLRNRLAA 3190

Qy       1358 SLKLKLSTTFLSTSPNIALLAQNLLDALATA 1388
               :    :|   :  :    |:  |  ||:  ||   :   |
Db       3191 ATGTRLPASAVFDHPHAAALARWLLAGMRHA 3221

RESULT 4
ERY3_SACER
ID   ERY3_SACER     STANDARD;      PRT;  3172 AA.
AC   Q03133; Q54097; Q99270;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE   deoxyerythronolide B synthase III) (DEBS 3).
GN   ERYA.
```

```
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=91043075; PubMed=2234082;
RA   Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT   "An unusually large multifunctional polypeptide in the erythromycin-
RT   producing polyketide synthase of Saccharopolyspora erythraea.";
RL   Nature 348:176-178(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: Contains 2 acyl carrier domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   PIR; S13595; S13595.
DR   PIR; S22012; S22012.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 5.
FT   DOMAIN     1485   3172       MODULE 6.
FT   DOMAIN       37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2926   3172       THIOESTERASE.
FT   ACT_SITE    199    199       THIOESTER BOND.
FT   ACT_SITE    643    643       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    1118   1164       NADP.
FT   BINDING    1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
FT   ACT_SITE   2112   2112       ACYL-ESTER INTERMEDIATE.
FT   NP_BIND    2557   2605       NADP.
FT   BINDING    2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    231    231       S -> SA (IN REF. 2).
FT   CONFLICT    240    240       MISSING (IN REF. 2).
FT   CONFLICT    289    289       A -> R (IN REF. 2).
FT   CONFLICT    493    493       P -> R (IN REF. 2).
FT   CONFLICT    493    517       PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT                                RWPPAAAPWEQ (IN REF. 1).
FT   CONFLICT    510    510       A -> R (IN REF. 2).
FT   CONFLICT    513    513       M -> W (IN REF. 2).
FT   CONFLICT    525    525       E -> D (IN REF. 2).
FT   CONFLICT    536    536       R -> G (IN REF. 2).
FT   CONFLICT    547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT    553    553       R -> G (IN REF. 2).
FT   CONFLICT    673    673       R -> A (IN REF. 2).
FT   CONFLICT    716    716       MISSING (IN REF. 2).
FT   CONFLICT    734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT    896    896       R -> RQR (IN REF. 2).
FT   CONFLICT    896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT    988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT   1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT   1124   1126       MISSING (IN REF. 1).
FT   CONFLICT   1132   1132       L -> V (IN REF. 2).
FT   CONFLICT   1192   1192       A -> R (IN REF. 2).
FT   CONFLICT   1194   1194       MISSING (IN REF. 2).
FT   CONFLICT   1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT   1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT   1485   1485       MISSING (IN REF. 2).
FT   CONFLICT   1518   1518       G -> R (IN REF. 2).
FT   CONFLICT   1601   1601       V -> L (IN REF. 2).
FT   CONFLICT   1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT   1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT   1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT   1762   1762       T -> S (IN REF. 2).
FT   CONFLICT   2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT   2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT   2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT   2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT   2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT   2596   2596       A -> G (IN REF. 2).
FT   CONFLICT   2609   2609       P -> A (IN REF. 2).
FT   CONFLICT   2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT   2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match             26.0%;  Score 1875;  DB 1;  Length 3172;
  Best Local Similarity   34.6%;  Pred. No. 6.9e-91;
  Matches  549;  Conservative  184;  Mismatches  602;  Indels  254;  Gaps   48;
```

```
Qy          4 RPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPD 63
Db         31 REVEHRAGEPIAIVGMACRFPGDVDSPESFWEFVSGGGDAIAEAPADR----GW-EPDPD 85

Qy         64 APGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALV 123
Db         86 A------RLGGMLAAAGDFDAGFFGISPREALAMDPQQRIMLEISWEALERAGHDPVSLR 139

Qy        124 GTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDT 183
Db        140 GSATGVFTGVGTVDYGPRPDE--APDEVLGYVGTGTASSVASGRVAYCLGLEGPAMTVDT 197

Qy        184 AYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEA 243
Db        198 ACSSGLTALHLAMESLRRDECGLALAGGVTVMSSPGAFTEFRSQGGLAADGRCKPFSKAA 257

Qy        244 DGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRA 303
Db        258 DGFGLAEGAGVLVLQRLSAARREGRPVLAVLAGSAVNQDGASNGLTAPSGPAQQRVIRRA 317

Qy        304 LADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEY 363
Db        318 LENAGVRAGDVDYVEAHGTGTRLGDPIEVHALLSTYGAERDPDDPLWIGSVKSNIGHTQA 377

Qy        364 ASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVS 423
Db        378 AAGVAGVMKAVLALRHGEMPRTLHFDEPSPQIEWDLGAVSVVSQARSWPAGERPRRAGVS 437

Qy        424 SFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQ 483
Db        438 SFGISGTNAHVIVEEAPEADEPEPAPDSGPVPLVLSGRDEQAMRAQAGRLADHLAPEPRN 497

Qy        484 CLGDVAFSLATTRSAMEHRLAVAATSRE---GLRAALDAAAQGQTSPGAVRSIADSSRGK 540
Db        498 SLRDTGFTLATRASAMEHRAVVVGDRDEALAGLRAVADRRIADRTATGQ----GPNSPRR 553

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVM-WAEPASVDAA 599
Db        554 VAMVFPGQGAQWQGMARDLLRESQVFADSIRDCERALAPHVDWSLTDLLSGARP------ 607

Qy        600 LLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVA 659
Db        608 -LDRVDVVQPALFAVMVSLAALWRSHGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVA 666

Qy        660 ARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAP-DQVVIAGAGQPVHAI 718
Db        667 VRSRVLRRLGGQGGMASFGLGTEQAAERIGRFAGALSIASVNGPRSVVVVAGESGPLDEL 726

Qy        719 AAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDE 778
Db        727 IAECEAEAHKARRIPVDYASHSPQVESLREELLTELAGISPVSADVALYSTTTGQPIDTA 786

Qy        779 VSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPA----- 833
Db        787 TMDTAYWYANLREQVRFQDATRQLAEAGFDAFVEVSPHPVLTVGIEATLDSALPADAGAC 846

Qy        834 LLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKA 893
Db        847 VVGTLRRDRGGLADFHTALGEAYAQGVEVDWSPAF-ADARPVELPVYPFQ--RYWLPIPT 903

Qy        894 DDAAR---GDRRAPGAGHDEVEEGGAVRGGDRRSARLDHP--PPE--SGRREKVEAAGDR 946
Db        904 GGRARDEDDDWRYQVVWREAEWESASLAG---RVLLVTGPGVPSELSDAIRSGLEQSGAT 960

Qy        947 PFRLEID---------EPGVLDHL----VLRVTERRA--PGLGEVEI--AVDAAG----- 984
Db        961 VLTCDVESRSTIGTALEAADTDALSTVGVAAVPHGEAVDPSLDALALVQALGAAGVEAPL 1020

Qy        985 --LSFNDVQLALGMVPDDLPGKPNPPLLGGECAGRIVAVGE-----GVNGLVVGQPVIA 1037
Db       1021 WVLTRNAVQVADGELVD--PAQ----AMVGG--LGRVVGIEQPGRWGGLVDLV------- 1065

Qy       1038 LSAGAFATHVTTSAALVL-PRPQALSAIEAAAMPVAYLTAWYALDRIARLQP-GERVLIH 1095
Db       1066 ---DADAASIRSLAAVLADPRGEEQVAIRADGIKVARLVPAPARARTHPLEPLAGTVLVT 1122

Qy       1096 AATGGVGLAAVQWAQHVGAEVHA---------TAGTPEKRAYLESLGVRY------VSDS 1140
Db       1123 GGTGGIGAHLARWLARSGAE-HLVLLGRRGADAPGASELREELTALGTGVTIAACDVADR 1181

Qy       1141 RSDRFVADVRAWTGGEGVDVVLN----SLSGELIDKSFNLLRSHGRFVELGKRDCYADNQ 1196
Db       1182 ARLEAVLAAEAAAEGRTVSAVMHAAGVSTSTPLDD------LTEAEFTEI------AD-- 1227

Qy       1197 LGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIA----AGVFTPPPIATLPIAR 1252
Db       1228 --------------VKVRGTV-----NLDELCPDLDAFVLFSSNAGVWGSPGLASYAAAN 1268

Qy       1253 V-ADAFRSMAQA-----------------------QHLGKLVLTLGDP-----EVQIRI 1282
Db       1269 AFLDGFARAARSEGAPVTSIAWGLWAGQNMAGDEGGEYLRSQGLRAMDPDRAVEELHITL 1328

Qy       1283 PTHAGAGPSTGDRD---------------LLDRLASA-----------APAARAAA---- 1312
Db       1329 -DHGQTSVSVVDMDRRRFVELFTAARHRPLFDEIAGARAEARQSEEGPALAQRLAALLCD 1387

Qy       1313 ------LEAFLRTQVSQVL-RTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLST 1365
Db       1388 GREREHLAHLIRAEVAAVLGHGDDAAIDRDRAFRDLGFDSMTAVDLRNRLAAVTGVREAA 1447

Qy       1366 TFLSTSPNIALLAQNLLDALATALSLERVAAE-------------------NLRAGVQN 1405
Db       1448 TVVFDHPTITRLADHYLERLVGAAEAEQAPALVREVPPKDADDPIAIVGMACRFPGGVHN 1507

Qy       1406 -----DFVSSG--------ADQDWEIIAL 1421
Db       1508 PGELWEFIVGGGDAVTEMPTDRGWDLDAL 1536
```

```
RESULT 5
MCAS_MYCBO
ID   MCAS_MYCBO     STANDARD;      PRT;  2110 AA.
AC   Q02251;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Mycocerosic acid synthase.
GN   MAS.
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1765;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=92406887; PubMed=1527058;
RA   Mathur M., Kolattukudy P.E.;
RT   "Molecular cloning and sequencing of the gene for mycocerosic acid
RT   synthase, a novel fatty acid elongating multifunctional enzyme, from
RT   Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL   J. Biol. Chem. 267:19388-19395(1992).
CC   -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
CC       METHYLAMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
CC       FORM MYCOCEROSYL LIPIDS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE.
CC   -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
CC       ARRANGEMENT.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: Contains 1 acyl carrier domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   -----------------------------------------------------------------------
DR   EMBL; M95808; AAA25369.1; -.
DR   PIR; B44110; B44110.
DR   HSSP; P73283; 1E5M.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00107; ADH_zinc_N; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW   Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT   DOMAIN        1    430       BETA-KETOACYL SYNTHASE.
FT   DOMAIN      533    852       ACYL TRANSFERASE.
FT   DOMAIN        ?      ?       ENOYL REDUCTASE.
FT   DOMAIN        ?      ?       BETA-KETOACYL REDUCTASE.
FT   DOMAIN     2026   2096       ACYL CARRIER (ACP).
FT   ACT_SITE    177    177       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    623    623       ACYL TRANSFERASES (BY SIMILARITY).
FT   NP_BIND    1561   1578       NADP (ER).
FT   NP_BIND    1765   1780       NADP (KR).
FT   BINDING    2059   2059       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   2110 AA;  225577 MW;  CC658215D7155300 CRC64;

  Query Match            25.7%;  Score 1852.5;  DB 1;  Length 2110;
  Best Local Similarity  29.5%;  Pred. No. 6.5e-90;
  Matches  524;  Conservative  176;  Mismatches  519;  Indels  555;  Gaps   30;

Qy         13 PIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKTPVTR 72
              |:|::|  |||||:         |  |     | |   :| :||||   ::||:|  ||::
Db          7 PVAVIGMGCRLPGGINSPDKLWESLLRGDDLVTEIPPDRWDADDYYDPEPGVPGRSVSRW 66

Qy         73 ASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGVFIG 132
                || ||| ||| ||||| |||  :||  |||||  |||:|:|  : |::|  |:  ||  |
Db         67 GGFLDDVAGFDAEFFGISEREATSIDPQQRLLLETSWEAIEHAGLDPASLAGSSTAVFTG 126

Qy        133 IGPSEYEAALPQATASAEIDAHGGL-------GTMPSVGAGRISYALGLRGPCVAVDTAY 185
              :    :|       |       |||          |  || :|||:: ||| ||  :  |||
Db        127 LTHEDYLVLTTTA---------GGLASPYVVTGLNNSVASGRIAHTLGLHGPAMTFDTAC 177

Qy        186 SSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADG 245
              || |:||||||:||  ||    |||||  :::|  |      |     |:    |||  :|  |:|||
Db        178 SSGLMAVHLACRSLHDGERDLALAGGCAVLLEPHACVAASAQGMLSSTGRCHSFDADADG 237

Qy        246 FGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALA 305
              | |  ||||:|:||||  |     ||:||  ||:||:|  | ||  :  ||:|:  :|   |  : |||
Db        238 FVRSEGCAMVLLKRLPDALRDGNRIFAVVRGTATNQDGRTETLTMPSEDAQVAVYRAALA 297

Qy        306 DAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYAS 365
              ||    :|| ||||||||  :|||||  ::|  |||           ||    :||  ||:|:||    ::
Db        298 AAGVQPETVGVVEAHGTGTPIGDPIEYRSLARVYGAG----TPCALGSAKSNMGHSTASA 353

Qy        366 GITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWN--TPRRAGVS 423
                ||:|:| :|||:||  :|   ||      |     :|       |   |  |  :|  ||||:  |   ||:|  ||
Db        354 GTVGLIKAILSLRHGVVPPLLHFNRLPDELSDVETGLFVPQAVTPWPNGNDHTPKRVAVS 413

Qy        424 SFGMSGTNAHVVLEEAPAATCTPPAPERPAE----LLVLSARTASALDAQAARLRDHLET 479
              |||||||| |  ::||||||       |  :    ||     |  :||: :: ||   |  :|   :|
Db        414 SFGMSGTNVHAIVEEAPAEASAPESSPGDAEVGPRLFMLSSTSSDALRQTARQLATWVEE 473

Qy        480 YPSQCL--GDVAFSLATTRSAMEHRLAVAATSR----EGLRAALDAAAQGQTSPGAVRSI 533
              :     |:    |:|::||  |      | || |  :     ||||     |     |   :  |
Db        474 H-QDCVAASDLAYTLARGRRHRPVRTAVVAANLPELVEGLREVADGDALYDAAVG----- 527

Qy        534 ADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEP 593
                    || :  ::|:|||:|    ||  |     |      |       :           |   |: ||
Db        528 -HGDRGPV-WVFSGQGSQWAAMGTQLL----ASEPVFAATI--------AKLEPVIAAES 573

Qy        594 A-SVDAALLDQTAFT-----QPALFTFEYALAA-LWRSWGVEPELVAGHSIGELVAACVA 646
                ||  |:   |     |     |||:|   :  ||||  :  :::||  |    |  |||:||   || ||
Db        574 GFSVTEAITAQQTVTGIDKVQPAVFAVQVALAATMEQTYGVRPGAVVGHSMGESAAAVVA 633

Qy        647 GVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADV-AAAVAPHAASVSIAAVNAPDQ 705
               |  ||||:  ::   |  :||    :    |||  |:|  |      |  :    :|      | :: |  :|
Db        634 GALSLEDAARVICRRSKLMTRIAGAGAMGSVELPAKQVNSELMARGIDDVVVSVVASPQS 693

Qy        706 VVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIV 765
               || |          |   :   |         |       :  :   |   | |||   :  |:|:           ::      |   :
Db        694 TVIGGTSDTVRDLIARWEQRDVMAREVAVDVASHSPQVDPILDDLAAALADIAPMTPKVP 753

Qy        766 LVSNLSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPA 825
                |              |     |||   :  |     |:||   |:|      |        |  :|  |        |        |
Db        754 YYSATLFDPREQPVCDGAYWVDNLRNTVQFAAAVQAAMEDGYRVFADVSPHPLLTHAVEQ 813

Qy        826 C--MPDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQ 883
                  |   || |   :   |   :   |    |       |   ::   |:|:  ||  |     |      |
Db        814 TGRSLDMSVAALAGMRREQPLPHGLRGLLTELHRAGAALDYSALYPA-GRLVDAPLPAWT 872

Qy        884 RERYWID---------------------------------------------------- 890
                |  :||
Db        873 HARLFIDDDGQEQRAQGACTITVHPLLGSHVRLTEEPERHVWQGDVGTSVLSWLSDHQVH 932

Qy        891 ------------------------------------------------------------ 890

Db        933 NVAALPGAAYCEMALAAAAEVFGEAAEVRDITFEQMLLLDEQTPIDAVASIDAPGVVNFT 992

Qy        891 -------------TKADDAARGDRRAPG-------------------------------- 905
                           |  |  ||   |    ||
Db        993 VETNRDGETTRHATAALRAAEDDCPPPGYDITALLQAHPHAVNGTAMRESFAERGVTLGA 1052

Qy        906 -----------------------------------------------------AGH-- 908
                                                                   ||
Db       1053 AFGGLTTAHTPRREPRRCWPRSRCPRRSGSSRAPTESTRRCWTLVSSRSARASSRYGHWW 1112

Qy        909 ----------------------------------------DEVEEGG----AVRG---- 919
                                                      |  ::|  |      ||||
Db       1113 PAVAVGCAQPACLRAYPQCPLLLHAVDQGLQRRDPRWEADLDVLDEHGTVLLAVRGLRMG 1172

Qy        920 -----------------------------GDRR---------SARLDHP---------- 930
                                            ||          |  :|  |
Db       1173 TGTSERDERDRLVSERLLTLGWQQRALPEVGDGEAGSWLLIDTSNAVDTPDMLASTLTDA 1232

Qy        931 ------------------------------------------------------------ 930

Db       1233 LKSHGPQGTECASLSWSVQDTPPNDQAGLEKLGSQLRGRDGVVIVYGPRVGDPDEHSLLA 1292

Qy        931 -------------------------------------PPESGRREKVEAAGDR-------- 946
                                                  | :||  |   :|  || |
Db       1293 GREQVRHLVRITRELAEFEGELPRLFVVTRQAQIVKPHDSGERANLEQAGLRGLLRVISS 1352

Qy        947 ------------------------------------------------------------ 946

Db       1353 EHPMLRTTLIDVDEHTDVERVAQQLLSGSEEDETAWRNGDWYVARLTPSPLGHEERRTAV 1412

Qy        947 ------PFRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVP-- 998
                    |:::   ||  |   |      ::|   || |:|:||    :  ::|  || :|  |  |
Db       1413 LDPDHDGMRVQVRRPGDLQTLEFVASDRVPPGPGQIEVAVSMSSINFADVLIAFGRFPII 1472

Qy        999 DDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALS-AGAFATHVTTSAALVLPR 1057
              ||    :|         ||  :    |  :|||||   ||     |      |    |  :  |   :|    |  |   :
Db       1473 DD--REPQ----LGMDFVGVVTAVGEGVTGHQVGDRVGGFSEGGCWRTFLTCDANLAVTL 1526

Qy       1058 PQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVH 1117
              |  |:   :|       |:  ||||  |:  :|:::  |::|||:||||||  ||:   |:   |||:
Db       1527 PPGLTDEQAITAATAHATAWYGLNDLAQIKAGDKVLIHSATGGVGQAAISIARAKGAEIF 1586

Qy       1118 ATAGTPEKRAYLESLGVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLL 1177
              ||||  |  |||  |    :||  :|  ||||   |       :|    |  |  |||:|||||:|                ||
Db       1587 ATAGNPAKRAMLRDMGVEHVYDSRSVEFAEQIRRDTDGYGVDIVLNSLTGAAQRAGLELL 1646
```

```
Qy       1178 RSHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAA 1237
                 |||||:|| | | : :||| || | |:|  :||  | : :| ||| ||   :   | |
Db       1647 PFGGRFVEIGKADVYGNTRLGLFPFRRGLTFYYLDLALMSVTQPDRVRELLATVFKLTAD 1706

Qy       1238 GVFTPPPIATLPIARVADAFRSMAQAQHLGKLVL 1271
              || | |      |:|  ||| |:|: |:| |||||
Db       1707 GVLTAPQCTHYPLAEAADAIRAMSNAEHTGKLVL 1740
```

RESULT 6

```
PPSB_MYCTU
ID   PPSB_MYCTU     STANDARD;      PRT;  1538 AA.
AC   Q10978; O53234;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsB.
GN   PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- SIMILARITY: Contains 1 acyl carrier domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL021070; CAA15929.1; -.
DR   EMBL; AE007122; AAK47329.1; -.
DR   PIR; E70874; E70874.
DR   TIGR; MT3002; -.
DR   TubercuList; Rv2932; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
KW   Transferase; NADP; Phosphopantetheine; Complete proteome.
FT   DOMAIN     1425   1495       ACYL CARRIER (ACP).
FT   BINDING    1458   1458       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   1538 AA;  162527 MW;  B55E2A2042AD00CC CRC64;

  Query Match             22.1%;  Score 1592.5;  DB 1;  Length 1538;
  Best Local Similarity   31.2%;  Pred. No. 2.2e-76;
  Matches  488;  Conservative  207;  Mismatches  600;  Indels  271;  Gaps   44;

Qy          3 DRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDP 62
              ||     |  :|:|:|| || || |       ||  |     |: :  |||:|||| |:: |||
Db         25 DRVSRIAVAEPVAVVGIGCRFPGDVDGPESFWDFLVAGRNAISTVPADRWDAEAFYHPDP 84

Qy         63 DAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSAL 122
                ||:          |: ||| |||  ||||:||||  |||   |:||||  |||||:| |  | :|
Db         85 LTPGRMTTKWGGFVPDVAGFDAEFFGITPREAAAMDPQQRMLLEVAWEALEHAGIPPDSL 144

Qy        123 VGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVD 182
               || |  | |:|:  :||::  |   | :   :||:  | |     |:  ||||| |||||| ||||
Db        145 GGTRTAVMMGVYFNEYQSML--AASPQNVDAYSGTGNAHSITVGRISYLLGLRGPAVAVD 202

Qy        183 TAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAE 242
              || |||||||||||||     |     ||||||: | | |   : :|      |:  ||| || |
Db        203 TACSSSLVAVHLACQSLRLRETDLALAGGVSITLRPETQIAISAWGLLSPQGRCAAFDAA 262

Qy        243 ADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKR 302
              |||| ||||  |||||||: |    |||::|||:||||:| || |:|:| || ::|   |:
Db        263 ADGFVRGEGAGVVVLKRLTDAVRDGDQVLAVVRGSAVNQDGRSNGVTAPNTAAQCDVIAD 322

Qy        303 ALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPE 362
              ||      |     ||  |||||||||| ||||| :|| | || | |          :|:||||:|| |
Db        323 ALRSGDVAPDSVNYVEAHGTGTVLGDPIEFEALAATYGHGGDACA---LGAVKTNIGHLE 379

Qy        363 YASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGV 422
              |:|| | :|  |::|   || :||    || |        |    |    :|||    |||| |
Db        380 AAAGIAGFIKATLAVQRATIPPNLHFSQWNPAIDAASTRFFVPTQNSPWPTAEGPRRAAV 439

Qy        423 SSFGMSGTNAHVVLEE-APAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYP 481
              ||||: ||||||::|: :  |   :      :     ||:: :||   : | |   | | :|
Db        440 SSFGLGGTNAHVIIEQGSELAPVSEGGEDTGVSTLVVTGKTAQRMAATAQVLADWMEGPG 499

Qy        482 SQ-CLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
              ::   : ||| ::   |:          | |   |     | | | | | :|| |  |   | | |
Db        500 AEVAVADVAHTVNHHRARQATFGTVVARDRAQAIAGLRALAAGQHAPGVV-SHQDGSPGP 558

Qy        541 -LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAA 599
               |:::|:|:|   ||||  |      |      :|  ::      ||:|:         :
Db        559 GTVFVYSGRGSQWAGMGRQLLADEPAFAAAVAELEPVFVEQAGFSLRDVI---ATGKELV 615

Qy        600 LLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVA 659
              ::|     |          :   ||||:||:|:||  |||:|| || ||| :  :  :  :|
Db        616 GIEQ---IQLGLIGMQLTLTELWRSYGVQPDLVIGHSMGEVAAAVVAGALTPAEGLRVTA 672

Qy        660 ARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIA 719
               | |||  |     | |   :      |     | :|  :    |::      |:| | ||||  :  :  :
Db        673 TRARLMAPLSGQGGMALLGLDAAATEALIADY-PQVTVGIYNSPRQTVIAGPTEQIDELI 731

Qy        720 AAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEV 779
              | : |:         :::  | |:| |  :   |             :: | |:| ::|        |  :
Db        732 ARVRAQNRFASRVNIEVAPHNPAMDALQPAMRSELADLTPRTPTIGIISTTYADLHTQPI 791

Qy        780 SSPGYWVRHAREVVRFADGVKALHAAGAG------TFVEVGPKSTLLGLVPACMPDA-RP 832
                 :|  :  |  |||      :|: :||:|       ||:|:         :   :  || ||
Db        792 FDAEHWATNMRNPVRFQ---QAIASAGSGADGAYHTFIEISAHPLLTQAIADTLEDAHRP 848

Qy        833 A-----LLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVP---------- 876
```

```
                  |:      ||   ||              |:  :    |
Db      849 TKSAAKYLSIGTLQRDADDTV-------------TFRTNLYTADIAHPPHTCHPPEPHPT 895

Qy      877 LPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEGGAVRGGDRRSARLDHP---PPE 933
             :|| |||   :|| |                                    ||    ||
Db      896 IPTTPWQHTHHWIAT-----------------------------------THPSTAAPE 919

Qy      934 SGRREKVEAAGDR----------------PFR--LEIDEPGV----------LDHLVLRV 965
                 ||   |                  | |   :  | |           | | : |
Db      920 DPGSNKVVVNGQSTSESRALEDWCHQLAWPIRPAVSADPPSTAAWLVVADNELCHELARA 979

Qy      966 TERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEG 1025
             : |    |    :|  :    :   |     ||    | : |    |  ||  | | ::
Db      980 ADSRVDSLSPPALAAGSDPAALLD---ALRGVDNVLYAPPVPGELLDIESAYQVFHATRR 1036

Qy     1026 VNGLVVGQPVIALSAGAFATHVTTSA---------------------ALVLPRPQALSAI 1064
             :   :|     |:|        :| :|                    :| |  |:    |
Db     1037 LAAAMVASSATAISPPKLFI-MTRNAQPISEGDRANPGHAVLWGLGRSLALEHPEIWGGI 1095

Qy     1065 ---------EAAAMPVAYLTAWYALD------------RIARLQ----PGERVLIHA--- 1096
                      | |   |  ||| :  |            : |||      ||: | ::|
Db     1096 IDLDDSMPAELAVRHV--LTAAHGTDGEDQVVYRSGARHVPRLQRRTLPGKPVTLNADAS 1153

Qy     1097 -----ATGGVGLAAVQWAQHVGAE-VHATAGTP----EKRAYLESLGVRYVS-------- 1138
                  ||| :|    ::    :||: :  | |  |       |  : |      ::
Db     1154 QLVIGATGNIGPHLIRQLARMGAKTIVAMARKPGALDELTQCLAATGTDLIAVAADATDP 1213

Qy     1139 ---DSRSDRFVADVRAWTG-----GEGVDVVLNSLSGELIDKSFN-----LLRSHGRFVE 1185
                :  |||  ::       |          |  :|: :: :  :      |      |  | | ::
Db     1214 AAMQTLFDRFGTELPPLEGIYLAAFAGRPALLSEMTDDDVTTMFRPKLDALALLHRRSLK 1273

Qy     1186 ---------------LGKR--------DCYADNQLGLRPFLR--------NLSFSLVDLR 1214
                           || |               :  |:  | |  :          |  || |::
Db     1274 SPVRHFVLFSSVSGLLGSRWLAHYTATSAFLDSFAGARRTMGLPATVVDWGLWKSLADVQ 1333

Qy     1215 --GMMLERPARVRALLEEL-LGLIAAGVFTPPPIATLPIAR----VADAFRSMAQAQHLG 1267
                 :   : :: : :|: :|  :     :      :||: :|      :| |:|:     : :
Db     1334 KDATQISAESGLQPMADEVAIGALPLVMNPDAAVATVVVAADWPLLAAAYRTRGALRIVD 1393

Qy     1268 KLVLT---LGDPEVQIRIPTHAGAGPSTGDRDLL-DRLASAAPAARAAALEAFLRTQVSQ 1323
              |:    :|  |  : |   |      : |:   ||:| |:: :  |
Db     1394 DLLPAPEDVGKGESEFR--TSLRSCPAEKRRDMLFDHVGALAATVMG------------- 1438

Qy     1324 VLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLD 1383
              |   :   |  | :||||||:|  |:  :   ||     |  : :    |  :    :| |
Db     1439 --MPPTEPLDPSAGFFQLGMDSLMSVTLQRALSESLGEFLPASVVFDYPTV----YSLTD 1492

Qy     1384 ALATAL 1389
             |||  |
Db     1493 YLATVL 1498
```

RESULT 7
PPSA_MYCTU

```
ID   PPSA_MYCTU     STANDARD;      PRT;  1876 AA.
AC   Q10977;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phenolpthiocerol synthesis polyketide synthase ppsA.
GN   PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
CC       SYNTHESIS.
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC       similarity).
CC   -!- SIMILARITY: Contains 2 acyl carrier domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z74697; CAA98988.1; -.
DR   EMBL; AE007122; AAK47328.1; -.
DR   PIR; C70749; C70749.
DR   TIGR; MT3000; -.
DR   TubercuList; Rv2931; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
KW   Phosphopantetheine; Complete proteome.
FT   DOMAIN        7     80       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1764   1833       ACYL CARRIER (ACP) 2.
FT   NP_BIND    1498   1503       NADP (POTENTIAL).
FT   BINDING      62     62       PHOSPHOPANTETHEINE (POTENTIAL).
FT   ACT_SITE    273    273       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    720    720       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1796   1796       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    624    624       D -> E (IN REF. 2).
FT   CONFLICT    877    877       R -> H (IN REF. 2).
FT   CONFLICT   1323   1323       G -> S (IN REF. 2).
SQ   SEQUENCE   1876 AA;  198834 MW;  D9783DBD48792110 CRC64;

  Query Match             22.0%;  Score 1583;  DB 1;  Length 1876;
  Best Local Similarity   29.2%;  Pred. No. 8.9e-76;
  Matches  532;  Conservative  205;  Mismatches  622;  Indels  460;  Gaps   49;

Qy          2 ADRPIERAA----EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAW 57
               :|  ::| |      ::|||:||  || |||:          |   |    | :: :|| :||
Db         88 SDAAVKRGARNSLDEPIAVVGMGCRFPGGISCPEALWDFLCERRSSISQVPPQRWQPFEG 147

Qy         58 FDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAI 117
                |:  |           ||| |:  ||| || |||  || :|||  ||||| |||||:| |
Db        148 GPPEVAAALARTTRWGSFLPDIDAFDAEFFEISPSEADKMDPQQRLLLEVAWEALEHAGI 207
```

```
Qy        118 APSALVGTETGVFIGIGPSEYEAALPQATAS-AEIDAHGGLGTMPSVGAGRISYALGLRG 176
              | |  : ||||  |   |||  |      |:|   :::|      |  |:  | |:|| |||
Db        208 PPGTLRRSATGVFAGACLSEYGA---MASADLSQVDGWSNSGGAMSIIANRLSYFLDLRG 264

Qy        177 PCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRC 236
              | |||||| |||||:|||||||: :|  |:| ||:|:|||:        :  ||:   |:|
Db        265 PSVAVDTACSSSLVAIHLACQSLRTQDCHLAIAAGVNLLLSPAVFRGFDQVGALSPTGQC 324

Qy        237 KAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQ 296
              :|| | |||| ||||  |||||||: |: ||||:|||| |||:| || |:||  || ::|
Db        325 RAFDATADGFVRGEGAGVVVLKRLTDAQRDGDRVLAVICGSAVNQDGRSNGLMAPNPAAQ 384

Qy        297 EIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKT 356
               ||:  |  :||    | | |||||||||| ||||||  :||    |  | ||   :||||||||
Db        385 MAVLRAAYTNAGMQPSEVDYVEAHGTGTLLGDPIEARALGTVLGRGRPEDSPLLIGSVKT 444

Qy        357 NLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNT 416
              |||| | |:|| |:| ||::||||||  : |  :   || |  ||||: |    :| ||
Db        445 NLGHTEAAAGIAGFIKTVLAVQHGQIPPNQHFETANPHIPFTDLRMKVVDTQTEWPATGH 504

Qy        417 PRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPERPA-ELLVLSARTASALDAQAARLRD 475
              |||||||||  |||||||:|:      |      ||   ||:: :|   :  | |  |  |
Db        505 PRRAGVSSFGFGGTNAHVVIEQGQEVRPAPGQGLSPAVSTLVVAGKTMQRVSATAGMLAD 564

Qy        476 HLETYPSQ-CLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIA 534
              :|    :   | |||| :|  ||      | |  |    | | | | ||  :|| |    |
Db        565 WMEGPGADVALADVAHTLNHHRSRQPKFGTVVARDRTQAIAGLRALAAGQHAPGVVNP-A 623

Qy        535 DSSRGK-LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEP 593
              | |  |    |:::|:|:|  ||||  |         ||  |      :|  ::     | :|:
Db        624 DGSPGPGTVFVYSGRGSQWAGMGRQLLADEPAFAAAVAELEPVFVEQAGFSLHDVL---- 679

Qy        594 ASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLED 653
                :    |    |    :  ||   ||  |:|| |:|| ||||:||: || |||    :   :
Db        680 --ANGEELVGIEQIQLGLIGMQLALTELWCSYGVRPDLVIGHSMGEVAAAVVAGALTPAE 737

Qy        654 AVFLVAARGRLMQALPAGG--AMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGA 711
               : : | | |||  |   |  |:: ::|| :    |   |    |::   |:| | ||||
Db        738 GLRVTATRSRLMAPLSGQGGMALLELDAPTTEALIADFP---QVTLGIYNSPRQTVIAGP 794

Qy        712 GQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLS 771
               : :  : | : |:       :::  | |:| |  :  |       :: | |:| ::|
Db        795 TEQIDELIARVRAQNRFASRVNIEVAPHNPAMDALQPAMRSELADLTPRTPTIGIISTTY 854

Qy        772 GKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAG------TFVEVGP----KSTLLG 821
                |  |   :|  :  |  |||   :|: :||:|·     ||:|:            ::
Db        855 ADLHTQPVFDAEHWATNMRNPVRFQ---QAIASAGSGADGAYHTFIEISAHPLLTQAIID 911

Qy        822 LVPACMPDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFP-----SGGRRVP 876
               : :  | ||  |  |  |  :  |   |                           |
Db        912 TLHSAQPGARYTSLGTLQRDTDDVVTFRTNLNKAHTI---------HPPHTPHPPEPHPP 962

Qy        877 LPTYPWQRERYWIDTKADDAARGDRRAPGAG-----HDEVEEGGAVRGGDRRSARL--DH 929
              :|| |||  |:|| ||     |     | ||     |     |       |||  |
Db        963 IPTTPWQHTRHWITTKYPAGSVGS--APRAGTLLGQHTTVATVSASPPSHLWQARLAPDA 1020

Qy        930 PPPESGRR----EKVEAA---------------------------GDRPFRLEIDEPGV 957
               | : | |    | | |:                            |||  :::   |
Db       1021 KPYQGGHRFHQVEVVPASVVLHTILSAATELGYSALSEVRFEQPIFADRPRLIQV----V 1076

Qy        958 LDHLVLRVTERRA--------------------------------------------P 971
               |:  :  :    |                                          |
Db       1077 ADNRAISLASSPAAGTPSDRWTRHVTAQLSSSPSDSASSLNEHHRANGQPPERAHRDLIP 1136

Qy        972 GLGEVEIAVDAAGLSF----------NDVQLALGMVPDDLPGKPNPPLLGGECAGRIV 1020
               ||:         ||  ||                  ::: :|: : |: ||     |||   :  |  :
Db       1137 DLAELLAMRGIDGLPFSWTVASWTQHSSNLTVAIDL-PEALPEGSTGPLL---DAAVHLA 1192

Qy       1021 AVGEGVNG-LVVGQPVIALSAGAFATHVTTSAAL-----------VLPRPQALSAIEAAA 1068
              |: :  :  | |    : :| |    |  :| |                |      : : :
Db       1193 ALSDVADSRLYVPASIEQISLGDVVTGPRSSVTLNRTAHDDDGITVDVTVAAHGEVPSLS 1252

Qy       1069 M-PVAYLTAWYALDRIARLQPGERVLIHAATGGVG--------LAAVQW-------AQHV 1112
               | :  |    : || : | ||         |:|| |           :   :  |       | |
Db       1253 MRSLRYRALDFGLD-VGRAQP------PASTGPVEAYCDATNFVHTIDWQPQTVPDATHP 1305

Qy       1113 GAE--VH-----------ATAGTPEKRAYLESLGVRYVSDSRSDRFVAD---------- 1148
              |||   |            |   |   |   |  ::     || :|   :|||
Db       1306 GAEQVTHPGPVAIIGDDGAALCETLEGAGYQPAVMSDGVSQARYVVYVADSDPAGADETD 1365

Qy       1149 ----VRAWTGGEG----------------------------------------------- 1157
                   || |  |
Db       1366 VDFAVRICTEITGLVRTLAERDADKPAALWILTRGVHESVAPSALRQSFLWGLAGVIAAE 1425

Qy       1158 --------VDVVLNSLSGELIDKSFNLLRSHGRFVELGKRD------------------ 1190
                      ||: :|   ||       ||     : : | :||
Db       1426 HPELWGGLVDLAINDDLGEFGPALAELLAKPSKSI-LVRRDGVVLAPALAPVRGEPARKS 1484

Qy       1191 --CYAD---------NQLGL--------------------RPFLRNLSFSLVD------ 1212
                 |  |          |||                     |  |:    :|
Db       1485 LQCRPDAAYLITGGLGALGLLMADWLADRGAHRLVLTGRTPLPPRRDWQLDTLDTELRRR 1544

Qy       1213 --------LRGMMLERPAR-------VRALL--------EELLGLI-------------- 1235
                      :||: :|   |        |:|||          : |:|
Db       1545 IDAIRALEMRGVTVEAVAADVGCREDVQALLAARDRDGAAPIRGIIHAAGITNDQLVTSM 1604

Qy       1236 -----------------------------------AAGVFTPPP----------IATLP 1249
                                                 |||:|  |           :  |
Db       1605 TGDAVRQVMWPKIGGSQVLHDAFPPGSVDFFYLTASAAGIFGIPGQGSYAAANSYLDALA 1664

Qy       1250 IARVADAFRSM-------------AQAQHLGKLVLTLGDPE------------------ 1277
               ||       :|              | || : : :  :| :
Db       1665 RARRQQGCHTMSLDWVAWRGLGLAADAQLVSEELARMGSRDITPSEAFTAWEFVDGYDVA 1724

Qy       1278 --VQIRIPTHAGAGPSTGDRDLLDRLASAAPAAR--AAALEAFLRTQVSQVLRTPEIKVG 1333
                | : :|  |||  |  :  ||       :  ||      :  ||  ||  ::  || || ::
Db       1725 QAVVVPMPAPAGADGSGANAYLLPARNWSVMAATEVRSELEQGLRRIIAAELRVPEKELD 1784

Qy       1334 AEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATALSLER 1393
                :  |  ||::||||: :|   |   : ::||  |    |  :  ||   |          :|
Db       1785 TDRPFFAELGLNSLMAMAIRREAEQFVGIELSATMLFNHPTVKSLASYL---------AKR 1835

Qy       1394 VAAENLRAGVQNDFVSSGA 1412
               ||  ::    |     :|| |
Db       1836 VAPHDVSQDNQISALSSSA 1854
```

RESULT 8
WA_EMENI
ID WA_EMENI STANDARD; PRT; 1986 AA.
AC Q03149;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Conidial green pigment synthase (EC 2.3.1.-).
GN WA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93101122; PubMed=1465094;
RA Mayorga M.E., Timberlake W.E.;
RT "The developmentally regulated Aspergillus nidulans wA gene encodes a
RT polypeptide homologous to polyketide and fatty acid synthases.";
RL Mol. Gen. Genet. 235:205-212(1992).
CC -!- FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN
CC INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED
CC BY CONIDIAL LACCASE TO FORM THE GREEN PIGMENT IN MATURE
CC ASEXUAL SPORES (CONIDIA).
CC -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC (POTENTIAL).
CC -!- PATHWAY: Conidial green pigment biosynthesis.
CC -!- SIMILARITY: Contains 2 acyl carrier domains.

```
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X65866; CAA46695.1; -.
DR   PIR; S28353; S28353.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Phosphopantetheine; Multifunctional enzyme; Repeat.
FT   DOMAIN      529    582       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   DOMAIN      991   1024       ACYL/MALONYL TRANSFERASES
FT                                (BY SIMILARITY).
FT   DOMAIN     1650   1719       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1772   1841       ACYL CARRIER (ACP) 2.
FT   ACT_SITE    548    548       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE   1001   1001       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1682   1682       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1804   1804       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1986 AA;  216634 MW;  74EF0940FF40EE9A CRC64;

  Query Match             16.4%;  Score 1180;  DB 1;  Length 1986;
  Best Local Similarity   26.9%;  Pred. No. 1.6e-54;
  Matches  419;  Conservative  242;  Mismatches  635;  Indels  264;  Gaps   53;

Qy          8 RAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGK 67
Db        373 RAELSKIAIIGMSGRFP-EADSPQDFWNLLYKGLDVHRKVPEDRWDADAHVDLTGTATNT 431

Qy         68 TPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTE- 126
Db        432 SKVPYGCWIREPGLFDPRFFNMSPREALQADPAQRLALLTAYEALEGAGFVPDSTPSTQR 491

Qy        127 --TGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184
Db        492 DRVGIFYGMTSDDYR----EVNSGQDIDTYFIPGGNRAFTPGRINYYFKFSGPSVSVDTA 547

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244
Db        548 CSSSLAAIHLACNSIWRNDCDTAITGGVNILTNPDNHAGLDRGHFLSRTGNCNTFDDGAD 607

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304
Db        608 GYCRADGVGTVVLKRLEDALADNDPILGVINGAYTNHSAEAVSITRPHVGAQAFIFKKLL 667

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLG--RDVATPLLIGSVKTNLGHPE 362
Db        668 NEANVDPKNISYIEMHGTGTQAGDAVEMQSVLDVFAPDHRRGPGQSLHLGSAKSNIGHGE 727

Qy        363 YASGITGLLKVVLSLQHGQIPAHLHAQA-LNPRISWGDLRLTVTRARTPWPDWNTP---- 417
Db        728 SASGVTSLVKVLLMMKENMIPPHCGIKTKINHNFPTDLAQRNVHIALQP-TAWNRPSFGK 786

Qy        418 RRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE----RPAELLVLSARTASAL----DAQ 469
Db        787 RQIFLNNFSAAGGNTALLLEDGPVSD-----PEGEDKRRTHVITLSARSQTALQNNIDAL 841

Qy        470 AARLRDHLETY---PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQ-- 524
Db        842 CQYISEQEKTFGVKDSNALPSLAYTTTARRIHHPFRVTAIGSSFQEMRDSLIASSRKEFV 901

Qy        525 ----TSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDL--CVRLFN 578
Db        902 AVPAKTPG-----------IGFLFTGQGAQYAAMGKQLYEDCSHFRSAIEHLDCISQ-G 948

Qy        579 QELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIG 638
Db        949 QDLPSILPLVDGSLPLSELSPVVVQLGTT-----CVQMALSSFWASLGITPSFVLGHSLG 1003

Qy        639 ELVAACVAGVFSLEDAVFLVAARGRLM--QALPAGGAMVSIEAPEADVAAAVAPHAASVS 696
Db       1004 DFAAMNAAGVLSTSDTIYACGRRAQLLTERCQPGTHAMLAIKAPLVEVKQLL--NEKVHD 1061

Qy        697 IAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAES 756
Db       1062 MACINSPSETVISGPKSSIDELSRACSEKGLKSTILTVPYAFHSAQVEPILEDLEKALQG 1121

Qy        757 VSYRRPSIVLVSNLSGKACTD---EVSSPGYWVRHAREVVRFADGVKALHAAGAG----T 809
Db       1122 ITFNKPSVPFVSALLGEVITEAGSNILNAEYLVRHCRETVNFLSAFEAVRNAKLGGDQTL 1181

Qy        810 FVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGL-- 867
Db       1182 WLEVGPHTVCSGMVKATL-GPQTTTMASLRRDEDTWKVLSNSLSSLYLAGVDINWKQYHQ 1240

Qy        868 -FPSGGRRVPLPTYPWQRERYWID-------TKADDAARGDRRAPGAGHDEVEEGGAVRG 919
Db       1241 DFSSSHRVLPLPTYKWDLKNYWIPYRNNFCLTKGSSMS----AASASLQPTFLTTSAQRV 1296

Qy        920 GDRRSARL--------DHPPPESGR---REKVEAAGDRPFRLEIDEPGVLDHLVLRVTER 968
Db       1297 VESRDDGLTATVVVHNDIADPDLNRVIQGHKVNGAALCPSSLYADSAQTLAEYLI---EK 1353

Qy        969 RAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNG 1028
Db       1354 YKPEL-------KGSGLDVCNVTV------------PKPLIAKTGKEQFRISATANWVDK 1394

Qy       1029 LVVGQPVIALSAG---------------AFATHVTTSAALVLPRPQALSAIEAAAMP--- 1070
Db       1395 HVSVQVFSVTAEGKKLIDHAHCEVKLFDCMAADLEWKRGSYLVK-RSIELLENSAVKGDA 1453

Qy       1071 ------VAYLTAWYALDRIARLQPGERVLI----HAATGGVGLAAVQ--------WAQ-- 1110
Db       1454 HRLRRGMVYKLFSALVDYDENYQSIREVILDSEHHEATALVKFQAPQANFHRNPYWIDSF 1513

Qy       1111 -HV-GAEVHATAGTPEKRAYLESLGVRYVSDSRS--DRFVADV-------------RAWT 1153
Db       1514 GHLSGFIMNASDGTDSK----SQVFVNHGWDSMRCLKKFSADVTYRTYVRMQPWRDSIWA 1569

Qy       1154 GG----EGVDVV-------LNSLSGELIDKSF----------NLLRSHG--RFVELGKRD 1190
Db       1570 GNVYIFEGDDIIAVFGGVKFQALSRKILDIALPPAGLSKAQTSPIQSSAPQKPIETAKPT 1629

Qy       1191 CYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAG-----VFTPPPI 1245
Db       1630 SRPAPPVTMKSFVKKSAGPSVVVRAL----------NILASEVGLSESDMSDDLVFADYGV 1680

Qy       1246 ATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPT-------------HAGAGPST 1292
Db       1681 DSLLSLTVTGKYRE--------ELNLDM-DSSVFIEHPTVGDFKRFVTQLSPSVASDSSS 1731

Qy       1293 GDR--------DLLDRLAS-AAPAARAAALEAFLRTQVSQVLR------TPEIKVGAEAL 1337
Db       1732 TDRESEYSFNGDSCSGLSSPASPGTVSPPNEKVIQIHENGTMKEIRAIIADEIGVSADEI 1791

Qy       1338 -----FTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATALSLE 1392
Db       1792 KSDENLNELGMDSLLSLTVLGKIRESLDMDL---------PGEFFIENQTLDQIETALDLK 1843
```

```
RESULT 9
MSAS_PENPA
ID   MSAS_PENPA     STANDARD;      PRT;  1774 AA.
AC   P22367;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   6-methylsalicylic acid synthase (EC 2.3.1.165) (6-MSAS).
OS   Penicillium patulum (Penicillium griseofulvum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX   NCBI_TaxID=5078;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=DSM 62862;
RX   MEDLINE=91006137; PubMed=2209605;
RA   Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E.;
RT   "The multifunctional 6-methylsalicylic acid synthase gene of
RT   Penicillium patulum. Its gene structure relative to that of other
RT   polyketide synthases.";
RL   Eur. J. Biochem. 192:487-498(1990).
CC   -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC       It catalyzes a total of 11 steps by seven different component
CC       enzymes, in the biosynthesis of the antibiotic patulin.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH = 6-
CC       methylsalicylate + 4 CoA + 3 CO(2) + NADP(+).
CC   -!- PATHWAY: Patulin biosynthesis.
CC   -!- SUBUNIT: HOMOMULTIMER.
CC   -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC   -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC       FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC       THIOLASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X55776; CAA39295.1; -.
DR   PIR; S13178; S13178.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
KW   Multifunctional enzyme; Transferase; Antibiotic biosynthesis; NADP;
KW   Phosphopantetheine.
FT   DOMAIN      186    238       ACYLTRANSFERASE (AT).
FT   DOMAIN      642    676       ACETYL/MALONYL TRANSFERASES.
FT   DOMAIN     1403   1450       2-OXOACYL REDUCTASE.
FT   DOMAIN     1700   1769       ACYL CARRIER (ACP).
FT   NP_BIND    1419   1424       NADP (POTENTIAL).
FT   ACT_SITE    204    204       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    653    653       MALONYLTRANSFERASE (BY SIMILARITY).
FT   BINDING    1732   1732       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1774 AA;  190732 MW;  05ED5DD10863F938 CRC64;

  Query Match             16.3%;  Score 1177;  DB 1;  Length 1774;
  Best Local Similarity   25.6%;  Pred. No. 2e-54;
  Matches  458;  Conservative  210;  Mismatches  655;  Indels  466;  Gaps   56;

Qy         14 IAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKTPVTRA 73
Db         35 VAVVGMACRVAGGNHNPELLWQSLLSQKSAMGEIPPMRWEPYYRRDARNEKFLKNTTSRG 94

Qy         74 SFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGVFIGI 133
Db         95 YFLDRLEDFDCQFFGISPKEAEQMDPQQRVSLEVASEALEDAGIPAKSLSGSDTAVFWGV 154

Qy        134 GPSEYEAA----LPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSSL 189
Db        155 NSDDYSKLVLEDLP------NVEAWMGIGTAYCGVPNRISYHLNLMGPSTAVDAACASSL 208

Qy        190 VAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGRG 249
Db        209 VAIHHGVQAIRLGESKVAIVGGVNALCGPGLTRVLDKAGAISSDGSCKSFDDDAHGYARG 268

Qy        250 EGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAGC 309
Db        269 EGAGALVLKSLHRALLDHDNVLAVIKGSAVCQDGKTNGIMAPNSVAQQLAANNALSAANI 328

Qy        310 AASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITG 369
Db        329 DPHTVRYVEAHATSTPLGDPTEISAIASVYGADRPADDPCYIGSIKPNIGHLEAGAGVMG 388

Qy        370 LLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSG 429
Db        389 FIKAVLAIQKGVLPPQANLTKLNSRIDWKTAGVKVVQEATPWPESDPIRRAGVCSYGYGG 448

Qy        430 TNAHVVLEE-APAATCTP---PAPERPAELLVLSARTASALDAQAARLRDHLETY-PSQC 484
Db        449 TVSHAVIEEFSPILQPDPLGNGAVSGPG-LLLLSGPQEKRLALQAKTLRDWMTAEGKDHN 507

Qy        485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQG--QTSPGAVRSIADSSRGKLA 542
Db        508 LSDILTTLATRRDHHDYRAALVVDDYRDAEQVLQSLANGVDHTFTTQSRVLGSDISKDVV 567

Qy        543 FLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLD 602
Db        568 WVFSGHGAQWPDMGKQL-----IHNPVFFAAI--------QPLDELIQAEIGLSPIELLR 614

Qy        603 QTAF-----TQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFL 657
Db        615 TGDFESSDRVQILTYVMQIGLSALLQSNGITPQAVIGHSVGEIAASVVAGALSPAEGALI 674

Qy        658 VAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHA 717
Db        675 VTRRALLYRQVMGKGGMILVNLPSAETEEILGSRSDLV-VAIDSSPSSCVVAGDKELVAE 733

Qy        718 IAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRV-AESVSYRRPSIVLVSNLSGKACT 776
Db        734 TAEALKARGVKTFTVKSDIAFHSPTLNGLVDPLRDVLAETLSPVSPNVKLYSTALADPRG 793

Qy        777 DEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDAR----- 831
Db        794 QDLRDVEYWAGNMVNRVRLTSAVKAAVEDGYRLFLEVSTHPVVSHSINETLMDAGMEDFA 853

Qy        832 --PALLASSRAGRDEPAT--VLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERY 887
Db        854 VIPTLL------RKKPTEKHILHSIAQLHCRGAEVNWAAQMP-GRWATGVPTTTWMHKPI 906

Qy        888 WIDTKADDAARGDRRAP---GAGHDEVEEGG------AVRGGDR--RSARLDH---PPPE 933
Db        907 W---------RKIETAPLHTGLTHD-VEKHTLLGQRIPVPGTDTYVYTTRLDNDTKPFPG 956

Qy        934 S---GRREKVEAAG-----------------------DRPFRLEI------------- 952
Db        957 SHPLHGTEIVPAAGLINTFLKGTGGQMLQNVVLRVPVAINAPRSVQVVVQQDQVKVVSRL 1016

Qy        953 --DEPGVLDHLVLRVT------ERRAPGLGE----------------------VEIAV 980
Db       1017 IPSEPSQLDDDASWVTHTTAYWDRKVAGSEDRIDFAAVKSRLVTKLADNFSIDYLDKVGV 1076

Qy        981 DAAGLSF-------NDVQLALGM-----VPDDLP-----------------GK---PNP 1007
Db       1077 SAMGFPWAVTEHYRNDKEMLARVDVNPAISGDAPLPWDSSSWAPVLDAATSVGSTIFPTP 1136
```

```
Qy       1008 PLLL----------------------------------------GGECAGRIVAVG---- 1023
              | :                                          ||   :  |:
Db       1137 ALRMPAQIERVEVFTSQDPPKISWLYVQEASDSVPTSHVSVVSEAGEVLAKFTAMRFSEI 1196

Qy       1024 EGVNGL------VVGQPVIALSAGAFATHVTTSAALVLPRPQALS-AIEAAAMP------ 1070
              || |:         :| |  ||       |     :    ::|  | |  : |: ||::|
Db       1197 EGTPGVSGSMESLVHQ--IAWPPATPAEEPLSIETVILVSPDATTRALYAASLPTRVNSF 1254

Qy       1071 ----------------------VAYLT----------------AWYALDRIARLQPGE-- 1090
                                    ||:                 |  |: |   |
Db       1255 QFSSTQEFFSNASSLPLEKGTVVTYIPGEVASLAEVPAASESFTWNLLELIKFTVNGSLP 1314

Qy       1091 -RVLIHAATGGVGLAAVQWAQ-------HVGAEVHATAGT----------PEKRAYLESL 1132
               :|    |  | |      ||       | |  |   ||                   |::
Db       1315 IKVFTLTANIGEGQTPTALAQSPLYGLARVIASEHPDLGTLIDVEEPVIPLSTMRYIQGA 1374

Qy       1133 GVRYVSD--SRSDRFVADVR--------------------AWTGGEGV------------ 1158
                :  ::|  :|: || :  |                    ||| ||
Db       1375 DIIRINDGIARTSRFRSLPRNKLLPASEGPRLLPRPEGTYLITGGLGVLGLEVADFLVEK 1434

Qy       1159 -------------------DVVLNSLSGELIDKSFNLLRSHGRFV-----ELGKRDCYAD 1194
                                  |  |   |  :      || | |  |          :: | |
Db       1435 GARRLLLISRRALPPRRTWDQVSEDLQPTI--AKIRLLESRGASVHVLPLDITKPDAVEQ 1492

Qy       1195 NQLGLRPFLRNLSF----SLVDLRG-------MMLERPARVRALLEELLGLIAAGVFTPP 1243
                  |   |  ||         :|    |       |   |   |  |  |  :: | :|      ||
Db       1493 ----LTTALDRLSLPSVQGVVHAAGVLDNELVMQTTRDAFNRVLAPKIAGALALHEVFPP 1548

Qy       1244 P----------------------------IATLPI--ARVADAF---------------- 1257
                                           : ||    ||: ||
Db       1549 KSVDFFVMFSSCGNLVGFTGQASYGSGNAFLDTLATHRARLGDAAVSFQWTSWRGLGMGA 1608

Qy       1258 --------------------RSMAQAQHLGKL--------------------VLTLGDPE 1277
                                  : |  |||  |                     |  |  |
Db       1609 STDFINAELESKGITDVTRDEAFAAWQHLAKYDMDHGVVLRSRAFEDGEPIPVSILNDIA 1668

Qy       1278 VQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEAL 1337
              |: |: | :   |:        | : :: |   : | |: :|  |::||:       | ::|
Db       1669 VR-RVGTVSNTSPAAAGSS--DAVPTSGPELK-AYLDEKIRGCVAKVLQMTAEDVDSKAA 1724

Qy       1338 FTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALA 1386
                    ||:|:|:| | || ::: :||: :  |    : | :: ||     : ||
Db       1725 LADLGVDSVMTVTLRRQLQLTLKIAVPPTLTWSHPTVSHLAVWFAEKLA 1773

RESULT 10
PKS1_ASPPA
ID   PKS1_ASPPA     STANDARD;      PRT;  2109 AA.
AC   Q12053;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Aflatoxin biosynthesis polyketide synthase (PKS).
GN   PKSL1.
OS   Aspergillus parasiticus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5067;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2999;
RX   MEDLINE=96042102; PubMed=7592391;
RA   Feng G.H., Leonard T.J.;
RT   "Characterization of the polyketide synthase gene (pksL1) required
RT   for aflatoxin biosynthesis in Aspergillus parasiticus.";
RL   J. Bacteriol. 177:6246-6254(1995).
CC   -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC       AFLATOXIN FROM HEXANOYL COA AND SEVEN MALONATES.
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
CC       similarity).
CC   -!- PATHWAY: Aflatoxin biosynthesis; first step.
CC   -!- SIMILARITY: Contains 1 acyl carrier domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L42766; AAC41675.1; -.
DR   EMBL; L42765; AAC41674.1; -.
DR   PIR; T17490; T17490.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00698; Acyl_transf; 1.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 1.
DR   Pfam; PF00550; pp-binding; 1.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE_NEG.
KW   Transferase; Acyltransferase; Phosphopantetheine;
KW   Multifunctional enzyme.
FT   DOMAIN      374    805       BETA-KETOACYL SYNTHASE.
FT   DOMAIN        ?      ?       ACYL/MALONYL TRANSFERASES.
FT   DOMAIN     1714   1785       ACYL CARRIER (ACP).
FT   DOMAIN        ?   2109       THIOESTERASE.
FT   ACT_SITE    543    543       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE    993    993       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1746   1746       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1937   1937       THIOESTERASE (BY SIMILARITY).
SQ   SEQUENCE   2109 AA;  230715 MW;  CB701372A16D8551 CRC64;

  Query Match             16.3%;  Score 1173.5;  DB 1;  Length 2109;
  Best Local Similarity   27.2%;  Pred. No. 3.8e-54;
  Matches  409;  Conservative  222;  Mismatches  625;  Indels  247;  Gaps   54;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKT 68
              | : :|||  | | |       | || ||     |       ||  |||     ||   |   |
Db        369 AGKCKLAIVSMSGRFPESPTTES-FWDLLYKGLDVCKEVPRRRWDINTHVDPSGKARNKG 427

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTE-- 126
                     :|      ||  ||||||:|| :|||| |: |     :||:| |  : |        |:
Db        428 ATKWGCWLDFSGDFDPRFFGISPKEAPQMDPAQRMALMSTYEAMERAGLVPDTTPSTQRD 487

Qy        127 -TGVFIGIGPSEYEAALPQATASAEIDAH----GGLGTMPSVGAGRISYALGLRGPCVAV 181
                |||  |:  :::       :     :   ||  :        |   | :|      |||::       ||
Db        488 RIGVFHGVTSNDW----METNTAQNIDTYFITGGNRGFIP----GRINFCFEFAGPSYTN 539

Qy        182 DTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSA 241
              ||| |||| |:|||| ||  |:|  ||:|||  ::: :|        ||       |:|| | ||  :
Db        540 DTACSSSLAAIHLACNSLWRGDCDTAVAGGTNMIYTPDGHTGLDKGFFLSRTGNCKPYDD 599

Qy        242 EADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLK 301
              :|||: | ||     | :|||   | || |  ||  | ||    :   ||      | :| |:  :|    :
Db        600 KADGYCRAEGVGTVFIKRLEDALADNDPILGVILDAKTNHSAMSESMTRPHVGAQIDNMT 659

Qy        302 RALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYG---LGRDVATPLLIGSVKTNL 358
                 ||    |   :     |:| ||||| :|| :|:::: :|:         |       || :|| | |:
Db        660 AALNTTGLHPNDFSYIEMHGTGTQVGDAVEMESVLSVFAPSETARKADQPLFVGSAKANV 719

Qy        359 GHPEYASGITGLLKVVLSLQHGQIPAHL---HAQALNPRI-SWGDLRLTVTRARTPWPDW 414
              || |   ||:| |:||:: :|| || |              :|        |    : :       |||
Db        720 GHGEGVSGVTSLIKVLMMMQHDTIPPHCGIKPGSKINRNFPDLGARNVHIAFEPKPWPRT 779

Qy        415 NTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLR 474
              :||||    :::|   :| |   :::|:||               | : ::  |||    :::        ||
Db        780 HTPRRVLINNFSAAGGNTALIVEDAPERHWPTEKDPRSSHIVALSAHVGASMKTNLERLH 839

Qy        475 DHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQ---GQTSPGAVR 531
```

```
OM protein - protein search, using sw model

Run on:          October  2, 2003, 17:27:33 ; Search time 49.4261 Seconds
                                              (without alignments)
                                              2764.850 Million cell updates/sec

Title:           US-10-014-717-2
Perfect score:   7210
Sequence:        1 VADRPIERAAEDPIAIVGAS..........GVQNDFVSSGADQDWEIIAL 1421

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:      283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        PIR_76:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2321	32.2	4151	2	G70944	probable polyketid
2	2298.5	31.9	2103	2	G86925	probable polyketid
3	2211.5	30.7	2478	2	AH2140	polyketide synthas
4	2207	30.6	1602	2	H70984	probable polyketid
5	2136	29.6	2518	2	AI2140	polyketide synthas
6	2120	29.4	1587	2	AB2012	hypothetical prote
7	2093.5	29.0	3573	2	S23070	erythronolide synt
8	2063.5	28.6	3413	2	T17467	rifamycin polyketi
9	2052.5	28.5	3739	2	T17410	polyketide synthas
10	2051.5	28.5	4613	2	T17409	polyketide synthas
11	2025	28.1	2126	2	H70621	probable polyketid
12	2014.5	27.9	2188	2	A70984	probable polyketid
13	2013.5	27.9	6260	2	T30228	polyketide synthas
14	2002.5	27.8	4735	2	T17463	rifamycin polyketi
15	2000.5	27.7	10223	2	T30225	polyketide synthas
16	1983.5	27.5	2201	2	S73014	polyketide synthas
17	1976	27.4	8563	2	T30226	polyketide synthas
18	1965	27.3	6420	2	T30283	polyketide synthas
19	1955	27.1	2100	2	T03223	probable polyketid
20	1925.5	26.7	3491	2	T43231	probable 6-deoxyer
21	1908.5	26.5	2108	2	H70819	probable polyketid
22	1900.5	26.4	2124	2	T28658	polyketide synthas
23	1889	26.2	3519	2	S43048	polyketide synthas
24	1875	26.0	3172	2	S22012	erythronolide synt
25	1869	25.9	2126	2	E70522	probable polyketid
26	1855	25.7	2111	2	A70668	mycocerosate synth
27	1852.5	25.7	2110	2	B44110	mycocerosate synth
28	1831.5	25.4	5069	2	T17464	rifamycin polyketi
29	1827	25.3	2116	2	C86926	probable mycoceros
30	1812	25.1	1774	2	T17421	polyketide synthas
31	1809	25.1	7576	2	T17428	FK506 polyketide s
32	1800	25.0	3178	2	S13595	6-deoxyerythronoli
33	1798	24.9	1562	2	T17411	polyketide synthas
34	1790.5	24.8	1402	2	D70634	probable polyketid
35	1785.5	24.8	1763	2	T17465	rifamycin polyketi
36	1784	24.7	1570	2	AC2012	hypothetical prote
37	1776.5	24.6	1762	2	T03222	probable polyketid
38	1773	24.6	1728	2	T17466	rifamycin polyketi
39	1751	24.3	1937	2	T03224	probable polyketid
40	1719.5	23.8	1017	2	B70985	probable polyketid
41	1712.5	23.8	2118	2	S72705	mycocerosate synth
42	1684.5	23.4	1293	2	T30871	orsellinic acid sy
43	1678	23.3	2297	2	T34918	polyketide synthas
44	1643.5	22.8	1827	2	B70984	probable polyketid
45	1638	22.7	1346	2	T17412	polyketide synthas

ALIGNMENTS

```
RESULT 1
G70944
probable polyketide synthase Rv2048c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70944
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70944
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-4151 <COL>
A;Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PIDN:CAA17262.1; PID:g289678
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks12
C;Superfamily: Mycobacterium tuberculosis probable polyketide synthase Rv2048c; 3-oxoacy
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1963-2038/Domain: acyl carrier protein homology <ACP1>
F;2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3995-4066/Domain: acyl carrier protein homology <ACP2>

  Query Match             32.2%;  Score 2321;  DB 2;  Length 4151;
  Best Local Similarity   33.2%;  Pred. No. 6.7e-117;
  Matches  613;  Conservative  203;  Mismatches  535;  Indels  498;  Gaps   31;

Qy          6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDA 64
              :   :|||||||| :|| ||||      | :|   || :   ||:| || |  ::|||||
Db       2051 VRTTSEDPIAIVGMACRYPGGVNSPDDMWDMLIQGRDVLSEFPADRGWDLAGLYNPDPDA 2110

Qy         65 PGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124
               |        |:  |  || :|||: | ||| ||| ||:|||: ||||| | | |: | |
Db       2111 AGACYTRTGGFVDGVGDFDPAFFGVGPSEALAMDPQHRMLLELSWEALERAGIDPTGLRG 2170

Qy        125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184
              : |||| |:    |        |:  ::     | : || :||::| ||| || |:||||
Db       2171 SATGVFAGVMTQGY-----GMFAAEPVEGFRLTGQLSSVASGRVAYVLGLEGPAVSVDTA 2225

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244
               ||||||:|:|  |||||||  ||||||::  :|   |  |: | |: |||||||:| ||
Db       2226 CSSSLVALHMAVGSLRSGECDLALAGGVTVNATPDIFVEFSRWRGLSPDGRCKAFAAAAD 2285
```

```
Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304
Db       2286 GTGFSEGGGMLVLQRLSDARRLGHPVLAVVVGSAVNQDGASNGLTAPNGPSQQRVVRAAL 2345

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGR-DVATPLLIGSVKTNLGHPEY 363
Db       2346 ANAGLSAAEVDVVEGHGTGTTLGDPIEAQALLATYGQDRGEPGEPLWLGSVKSNMGHTQA 2405

Qy        364 ASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVS 423
Db       2406 AAGVAGVIKMVLAMRHELLPATLHVDVPSPHVDWSAGAVELLTAPRVWPAGARTRRAGVS 2465

Qy        424 SFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQ 483
Db       2466 SFGISGTNAHVIIEAVPVVPRREAGWAGPVVPWVVSAKSESALRGQAARLAAYVRGDDGL 2525

Qy        484 CLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAF 543
Db       2526 DVADVGWSLA-GRSVFEHRAVVVGGDRDRLLAGLDELAGDQLGGSVVRGTA-TAAGKTVF 2583

Qy        544 LFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQ 603
Db       2584 VFPGQGSQWLGMGMGLHAGYPVFAEAFNTVVGELDRHLLRPLREVMWGH----DENLLNS 2639

Qy        604 TAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGR 663
Db       2640 TEFAQPALFAVEVALFRLLGSWGVRPDFVMGHSIGELSAAHVAGVLSLENAAVLVAARGR 2699

Qy        664 LMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMA 723
Db       2700 LMQALPAGGAMVAVQAAEEEVRPLL---SAEVDIAAVNGPASLVISGAQNAVAAVADQLR 2756

Qy        724 ARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPG 783
Db       2757 ADGRRVHQLAVSHAFHSPLMDPMIDEFAAVAAGIAIGRPTIGVISNVTGQLAGDDFGSAA 2816

Qy        784 YWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRD 843
Db       2817 YWRRHIRQAVRFADSVRFAQAAGGSRFLEVGPSGGLVASIEESLPDVAVTTMSALRKDRP 2876

Qy        844 EPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI--DTKADDAA---- 897
Db       2877 EPATLTNAVAQGFVTGMDLDWRAVV-GEAQFVELPTYAFQRRRFWLSGDGVAADAAGLGL 2935

Qy        898 ----------------------------------------RGDRRAPGAGH--------DE 910
Db       2936 AASEHALLGAVIDLPASGGVVLTGRLSPSVQGWLADHSVAGVTIFPGAGFVELAIRAGDE 2995

Qy        911 V--------------------------------EEG---------------------GAV 917
Db       2996 VGCGVVDESTLAAPLVLPASGSVAVQVVVNGPDESGVRGVSVYSRGDVGTGWVLHAEGAL 3055

Qy        918 RGG--------------------------------------------------------- 920
Db       3056 RAGSAEPTADLAMWPPAGAVPVEVADGYQQLAERGYGYGPAFRGLTAMWRRGDEVFAEVA 3115

Qy        921 ------------------------------------------------------------ 920
Db       3116 LPADAGVSVTGFGVHPVLLDAALHAVVLSAESAERGQGSVLVPFSWQGVSLHAAGASAVR 3175

Qy        921 ---------------------------------------------------DR---------- 922
Db       3176 ARIAPVGPSAVSIELADGLGLPVLSVASMLARPVTDQQLRAAVSSSGPDRLFEVTWSPQP 3235

Qy        923 ------------------------------------------------------------ 922
Db       3236 SAAVEPLPVCAWGTTEDSAAVVFESVPLAGDVVAGVYAATSSVLDVLQSWLTRDGAGVLV 3295

Qy        923 --------------------------RSARLDHP------------------------P 931
Db       3296 VMTRGAVALPGEDVTDLAGAAVWGLVRSAQTEHPGRIVLVDSDAPLDDSALAAVVTTGEP 3355

Qy        932 PESGRREKVEAA---------------GDRPFRLEIDEPGVLDHLVLR-VTERRAP-GLG 974
Db       3356 QVLWRRGEVYTARVHGSRAVGGLLVPPSDRPWRLAMSTAGTFENLRLELIPDADAPLGPG 3415

Qy        975 EVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQP 1034
Db       3416 QVRVAVSAIAANFRDVMIALGLYPD-------PDAVMGVEACGVVIETSLNKGSFAVGDR 3468

Qy       1035 VIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLI 1094
Db       3469 VMGLFPEGTGTVASTDQRLLVKVPAGWSHTAAATTSVVFATAHYALVDLAAARSGQRVLI 3528

Qy       1095 HAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGV--RYVSDSRSDRFVADVRAW 1152
Db       3529 HAGTGGVGMAAVQLARHLGLEVFATA-SKGKWDTLRAMGFDDDHISDSRSLEFEDKFRAA 3587

Qy       1153 TGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR-PFLRNLSFSLV 1211
Db       3588 TGGRGFDVVLDSLAGEFVDASLRLVAPGGVFLEMGKTDIRDPGVIAQQYPGVRYRAFDLF 3647

Qy       1212 DLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLV- 1270
Db       3648 E------PGPDRIAQILAELATLFGDGVLRPLPVTTFDVRCAPAALRYLSQARHTGKVVM 3701

Qy       1271 -----------LTLGDPEVQIRIPTHAGA------------GP-STGDRDLLDRLASA 1304
Db       3702 LMPGSWAAGTVLITGGTGMAGSAVARHVVARHGVRNLVLVSRRGPDAPGAAELVAELAAA 3761

Qy       1305 APAARAAALEAFLRTQVSQVLRTPEIK---------VGA--EALFTRLGMDSLMAVELRN 1353
Db       3762 GAQVQVVACDAADRAALAKVIADIPVQHPLSGVIHTAGALDDAVVMSLTPDRVDVV-LRS 3820

Qy       1354 RIEASLKLKLSTTFLSTSPNIAL----------------LAQNLLDALA 1386
Db       3821 KVDAAWHLHELTRDLDVSAFVMFSSMAGLVGSSGQANYAAANSFLDALA 3869
```

RESULT 2
G86925
probable polyketide synthase (imported) - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 24-May-2001
C;Accession: G86925
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; H
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: G86925
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2103 <STO>
A;Cross-references: GB:AL450380; NID:g13092513; PIDN:CAC29643.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0135
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I h
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein

```
  Query Match             31.9%;  Score 2298.5;  DB 2;  Length 2103;
  Best Local Similarity   32.9%;  Pred. No. 4.5e-116;
  Matches  627;  Conservative  204;  Mismatches  536;  Indels  541;  Gaps   38;

Qy          4 RPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDP 62
Db         34 REYEQRATEPVAVVGIGCRFPGGVDSPDRLWDVVSECRDVVSEFPTDRGWDVDGLFDPDP 93

Qy         63 DAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSAL 122
Db         94 DAEGKTYTRWGAFLDDAAGFDAGFFGIAPSEVLAMDPQQRLMLEVSWEALEYAGIDPLSL 153
```

```
Qy        123 VGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVD 182
Db        154 RGSATGVFTGIFAPSY-----GSRDTGGLQGYGLTGTAVSVASGRVSYVLGLEGPAVSVD 208
Qy        183 TAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAE 242
Db        209 TACSSSLVAIHWAMSSLRSGECDLALAGGVTVMGLPSIFVGFSRQRGLAADGRCKAFARA 268
Qy        243 ADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKR 302
Db        269 ADGTGWGEGAGVVVLERLSDAQRLGHTVLAVVRGSAVNQDGASNGLTAPNGLAQQRVIQA 328
Qy        303 ALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPE 362
Db        329 ALTNAGLSSADVDVVEAHGTATTLGDPIEAQALLATYGQGRPADQPLWVGSIKSNMGHTQ 388
Qy        363 YASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTV-TRARTPWPDWNT---PR 418
Db        389 AAAGVAGVIKMVQAMRHHVMPATLHVDEPSPRVDWTTGAVSVLTEAR----DWSVEGRPR 444
Qy        419 RAGVSSFGMSGTNAHVVLEEAP----AATCTPPAPERPAELLVLSARTASALDAQAARLR 474
Db        445 RAGVSSFGISGTNAHVILEEAPTPVEASYSTADQQRLSVVPWVVSGRSTAALMAQASRLA 504
Qy        475 DHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIA 534
Db        505 FYVQADPEVDPVDVGCALA-ARSVFEHRAVVVGESREQLIAGLAGLAVGESGAGVAIGQA 563
Qy        535 DSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPA 594
Db        564 -APLGKTVVVFPGQGAQRIGMGRELCSQLPVFAEEFDAVTGELDRHMRLPLRDVVW---- 618
Qy        595 SVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDA 654
Db        619 GADAGLLDSTEFAQPALFAVEVALFAVLQHWGLQPDFVMGHSIGELAAAYVAGVLAPADA 678
Qy        655 VFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQP 714
Db        679 AMLVVARGRLMQSMPAGGTMWAVAASEDEVVSLL---GEGVGIAAINAPESVVISGPQAA 735
Qy        715 VHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKA 774
Db        736 VSAIADKFAAQGRRVHQLAVSHAFHSPLMEPMLEEFVRVAAQVEVRKPQIGLVSNVTGEL 795
Qy        775 CTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPAL 834
Db        796 AGPDFGSPQYWGEHVSRAVRFVDSARHLQTLGATHFIESGPSSGLMGAIEQSLAPAEAVV 855
Qy        835 LASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFP-SGGRRVPLPTYPWQRERYW----- 888
Db        856 VSMMGKDRPELASVLGAAGQLFATGMPVEWPAVFAGSSRRRVTLPTYAFQRRRFWETPGT 915
Qy        889 ------------------------------------------------------------ 888
Db        916 DGTASVSGLGLGSTEHALLGAVVERPDSGGVVLTGRLSLANQPWLADHVIGGVVLFPGAG 975
Qy        889 --------------------------------------IDTKADDA--------ARGDR- 901
Db        976 FVELVIRAGDEVGCAGVEELVLAAPLVLHPGSGVQVQVVVAAADESSRRAVSVYSRGDQS 1035
Qy        902 ------------------------------------------------------------ 901
Db       1036 HGGWLLNAEGLLEVDAAEATVDLSLWPPEGAESVDISDAYPQLAERGYAYGPAFQGLVAI 1095
Qy        902 ------------------------------------------------------------ 901
Db       1096 WRRGSELFAEVVAPTGVVVDGMGMHPALLDAVLHALGLAIENTQATTEMRLPFCWRGVSL 1155
Qy        902 ----------RAPGAGHDEV--------------------------------EEGGAV- 917
Db       1156 HAGGAGRVRARFTSAGADAIAVDIADSAGLPVLTVRSLVTRAMTAEQLRTAVTAAGGALE 1215
Qy        918 ----------------------------------RGGDR-------------------- 922

Db       1216 QGPMDLVWSPIPLSSDGHGGGLPSVVSWADYCAGRGNDTGVVVWEFGSASAGAQAPVDSV 1275
Qy        923 -----------------------------------------------------RSARLDH 929
Db       1276 AAVYSATHAALQVLQFWFGQDRAGTLVVLTRGAVGLVGEGISDLAAAAVWGMVRSAQSEH 1335
Qy        930 PPPESG-------------------------RREKVEAA-------------GDRPPRL 950
Db       1336 ----SGKIVLIDTDSAVDVAVLADLGESQLLVRCNTVHAARLSPAPQLLALPVGESMWRL 1391
Qy        951 EIDEPGVLDHLVLR-VTERRAP-GLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPP 1008
Db       1392 AADGGGTLEDLVIQPCPEVQAPLQAGHMRVAVAAVGVNFRDVVAALGMYPGQAPP----- 1446
Qy       1009 LLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAA 1068
Db       1447 --LGAEGAGVVIEIGPEVTGVAVGDAVMGFLGGAGPLAV-VHQQLITQMPQGWSLAEAAA 1503
Qy       1069 MPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAY 1128
Db       1504 VPVVFLTALFGLADLAGIRAGESVLIHAGTGGVGMAAVQFARHWGVEIFVTA-SRGKWDT 1562
Qy       1129 LESLGV--RYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVEL 1186
Db       1563 LRDMGFDDDHIGDSRTLEFEEKFLAVTDGRGVDVVLDSLAGDFVDASLRLLVRGGRFLEM 1622
Qy       1187 GKRDCYADNQLGLR-PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPI 1245
Db       1623 GKTDIRDAQEIAANYPGVQYRAFDLSE------AGPVRMQEMLCEVRELFDTQVLHRLPV 1676
Qy       1246 ATLPIARVADAFRSMAQAQHLGKLVLTL------GDPEVQIRIPTHAGA----------- 1288
Db       1677 TSWDVRCAPAAFRFMSQARHIGKVVLTMPSALADGLAEGTVLITGASGAIGGVLARHMVS 1736
Qy       1289 ---------GPSTGDR-----DLLDRLASAAPAARAAALEA----FLRTQVSQVLRT-PE 1329
Db       1737 AYGVRHLVLASRRGDRAEGAVELAAELVGAGAKIQVVACDVADNDAVRGLFAQLSREFPP 1796
Qy       1330 IK-------VGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL 1382
Db       1797 VRGVIHAAGVLDDGAITSLTPDRVDTV-LRAKVDAAWNLHEATR------DLDLLMFVLC 1849
Qy       1383 DALATALS-------------LERVAAENLRAGVQNDFVSSGADQDWE 1417
Db       1850 SSIAATVGSPGQGNYSAANAFLDGLAAYRQAAGLAGTSLAWGL---WE 1894
```

RESULT 3
AH2140
polyketide synthase [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AH2140
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2140
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2478 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB74378.1; PID:g17131772; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr2679

```
  Query Match             30.7%;  Score 2211.5;  DB 2;  Length 2478;
  Best Local Similarity   25.1%;  Pred. No. 2.8e-111;
  Matches  609;  Conservative  242;  Mismatches  502;  Indels 1075;  Gaps   28;
```

```
Qy      6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65
Db     51 LEAQQTEPIAIIGMACRFPGGANNPEAFWQLLRDGVNAVGEIPDARWDIEKYYDANPDTP 110
Qy     66 GKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGT 125
Db    111 GKMYTRYGHFLDKVDEFDAQFFGISPREAQSLDPQQRLLLEVSYEALERAGQSPAKIKGS 170
Qy    126 ETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAY 185
Db    171 QTSIYIGLCFDDY--AKHSLSHPTQIDAFSSLGNTRSIAAGRIAYVLGLQGTVMQLDTTC 228
Qy    186 SSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADG 245
Db    229 SSSLLGVHLACQSLRNGESDMALAGGVNLILSPEPMIGFCKLKALAVDGRCKTLDAAADG 288
Qy    246 FGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALA 305
Db    289 YGRGEGCGIVVLKKLRDAIANRDPILAVIRGSAVNHDGSSNGLTAPNGTAQEAVIHQALQ 348
Qy    306 DAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYAS 365
Db    349 NAQVEPHQIQYSELHGTGTVLGDPIEVLALAKVLGEGRSKNQPLSIGSVKTNIGHLEGAA 408
Qy    366 GITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSF 425
Db    409 GVAGLMKVILALQHQQIPAHVNFQQPNPYIPWQKLPLTVPTQLTPWQSANSQRLAGVSSF 468
Qy    426 GMSGTNAHVVLEEAPAATCTPPAP----ERPAELLVLSARTASALDAQAARLRDHLETYP 481
Db    469 GMSGTNVHLILEE------FTPVPNSQQDTGYQLLTISAKTPQALRELVGEYIKYFASQP 522
Qy    482 SQCLGDVAFSLATTRSAMEHRLAVAATS----REGLRAALDAAAQGQTSPGAVRSIADSS 537
Db    523 EDNLDDICFTSNVGRSHFPYRLTIPANSYNSVYEQLSNYLNNTAKTFDS-----QLIPPN 577
Qy    538 RGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWA--EPAS 595
Db    578 TLKIAFLFTGQGSQYLNMARELYDTHPYFRQQVDKCCELLHPYLDVDLREILFTPETPNS 637
Qy    596 VDAAL------LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVF 649
Db    638 SNYELRITNYELSHTIYTQPALFVIEYALTQLWRHWGVTPDFLIGHSLGEYVAACIAGVF 697
Qy    650 SLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIA 709
Db    698 TLPDALKLIAHRARLMQQLPENGKMLAVSASATEIAEYIQAYSENVGIAAINSPENTVIS 757
Qy    710 GAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSN 769
Db    758 GEQTAIAQLAEILQNQGIKTTPLSVSHAFHSPLMETMVEEFRQVAESITFHPPKTPIISN 817
Qy    770 LSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPD 829
Db    818 ITGTVIGEEIANAEYWCHHILQPVQFALGIETLAKDNCKIFLEIGVKPTLLSIGKTCFPD 877
Qy    830 ARP-ALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPS-GGRRVPLPTYPWQRERY 887
Db    878 AENYSWLPSLRPRQSDRQTILSSLASLYMQGIDINWENFTDNQPHQRVILPTYSFQRQRF 937
Qy    888 WID--------------------------------------------------------- 890
Db    938 WVDKPEFTPPYPTPNLHPLLGQKLKLAKSEKIYYQNKISPKKPDFLQQHQVQSNSIMPAA 997
Qy    891 ------------------------------------------------------------ 890
Db    998 GYIEIAITAAEQIFPSHSYNLQNIQIIQPLPTTQEITLQLILTPLEKQTYQFEILSLCAF 1057
Qy    891 --------------------------------TKADDAA--------------------- 897
Db   1058 APLREKETTHATGIIQTSPPLPNSPQLPTTCTQKIDPTTYYQTCQKLGIEYGKYFQAIQQ 1117
Qy    898 --RGDRRAPG-------------------------------------------------- 905

Db   1118 LWRGENQALGEIKLSPEIDTTPYQIHPIILDACLQVTGAIFLDSQPLTAYLPIAIEQFQW 1177
Qy    906 ------------------------------------------------------------ 905
Db   1178 YGEKHQTHLWSYVKLRERYIADIQIFNANNKLIATITGLKIQPVKQLNNWQDWLYSIAWK 1237
Qy    906 -------------------------------------------------AGHD------- 909
Db   1238 ENPLTYHHLLTPTKVSNSSILNQSELDNYLSLSSQLEKLSIACILQTLQAGEELSIVTQH 1297
Qy    910 ------------------------------------------------EVEEG------- 914
Db   1298 KKLYQYLLQILQNQEQRETPLSFTQLRQTYPNATSELNLLKRCTSHLAEVLQGKIDPLQL 1357
Qy    915 -------------------------------------------------GAVRGG----- 920
Db   1358 LFPNGDFSDLTHLYQNSPIAKMMNTLVKQAITTAITDIPPEQPLRILEIGAGTGGTTAHL 1417
Qy    921 -----------------------------DRRSAR-----LDHPPPESGRREK------- 939
Db   1418 LPELAQRNIEYFFTDVSPLFLSKAQQQFADYSCVRYQLFDVEKPPTTQGFTENSFDIIIA 1477
Qy    940 ------------------------------------------------------------ 939
Db   1478 ANVLHATEDLRQTIIHIKQLLSRGGLLILLEGVQPVWWLDLIFGLTEGWWRFQDYDLRPH 1537
Qy    940 -----------------------VEAAGDRPF---------------------------- 948
Db   1538 HPLISSEQWQKLFTQEGFDSTTLLETIGNQAVITAQTPKLPDNCVILADKQGLGEKLATH 1597
Qy    949 ------------------------------------------------------------ 948
Db   1598 LQTQGVTCHLIFADSQPPTLPTPASHLINLWGLDVPHPDELTTAELIKQQQILCESLLYS 1657
Qy    949 ------------------------------------------------------------ 948
Db   1658 LQNNIKYQQLPQWWLVTQGAVNTDQDTSIKGLAATTLWGISKVIQLEHPEINCHCLDLDP 1717
Qy    949 --------------------------------------------------------RLEI 952
Db   1718 NHTPIEQLDILLKELLTNSPEKEIAWRTNIRYTPRLSQNSLCVSAPLREKNNSLMQRLTI 1777
Qy    953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012
Db   1778 PQRGTLENLTFQTVNRTSPNPGEIEIRIQATGLNFRDILNAL----DLYPGEPG---MLG 1830
Qy   1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAA-------LVLPRPQALSAIE 1065
Db   1831 CECVGEIVAIGTDVKHLQIGQTVIALASGSFSQYVTIKAAMASPAAGIAIPLPQNLNIID 1890
Qy   1066 AAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEK 1125
Db   1891 GATIPAAFLTAFYTLRHLAKIRPGDKVLIHAAAGGVGQAAIQIAKLAGAEIFATA-SPQK 1949
Qy   1126 RAYLESLGVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVE 1185
Db   1950 WETLRNLGVTKIFNSRTLDFAEEILTQTQGEGVDIVLNSLRGDFIAASFAVLKPQGRFVE 2009
Qy   1186 LGKRDCYADNQL-GLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPP 1244
Db   2010 IGKIDVWTEEQVKQVKP---NANYFLVDLVDLCQQQPELIQNMLKQLVEEFEQHRLQPLA 2066
Qy   1245 IATLPIARVADAFRSMAQAQHLGKLVLTLG------------------------------ 1274
Db   2067 AKVFPMNQIVDAFRYMQQGKHIGKVVVDWGLGNSLYPLKENCTYLITGGLGGLGLLVAEW 2126
Qy   1275 ------------------------------------------------------------ 1274
Db   2127 LVKMGACHLLLLGRNAPTADAVSKIQELETTGARIAIAQIDVSQTAALSNILTQIENSAY 2186
Qy   1275 ------------------------------------------------------------ 1274
```

```
Db     2187 PLAGIIHAAGVLDDGALLQMNWERFQKVMAAKVMGAWNLHTLTQHLPLDFFVLFSSATAL 2246

Qy     1275 ------------------------------------------------------------ 1274

Db     2247 FGSPGQANHVAANTFLDTLAHYRHTKGLPAMSINWGIWSEIGSATQATKQMQQRGIEAIA 2306

Qy     1275 ------------DPEVQIRI-------------------PTHAGAGPSTGD-----RDL 1297

Db     2307 PADGIQILQHLMTQPLTQIGVVPINWQKFPINSPFFENFYPTN---NPETRQTPPPKTDL 2363

Qy     1298 LDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEA 1357

Db     2364 LEKLSQLDKQTGYELLETYLQEQIAKVLGFPPHEINPQTGFFDLGMDSLTAIEFKNRLQT 2423

Qy     1358 SLKLKLSTTFLSTSPNIALLAQNLLDAL 1385

Db     2424 DLKITLPSTIAFDYPNIQTLAHHLNNQL 2451
```

RESULT 4
H70984
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: H70984
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70984
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1602 <COL>
A;Cross-references: GB:Z95617; GB:AL123456; NID:g3242249; PIDN:CAB09098.1; PID:e317262;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks8
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
F;55-452/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;561-839/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>

```
  Query Match             30.6%;  Score 2207;  DB 2;  Length 1602;
  Best Local Similarity   35.0%;  Pred. No. 2.8e-111;
  Matches  555;  Conservative  141;  Mismatches  460;  Indels  428;  Gaps   17;

Qy        6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDA 64

Db       28 LEAKLSEPVAVVGMGCRYPGGVDSPETLWELVAQGRDAVSDFPADRGWDVDGLFDPDPDA 87

Qy       65 PGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124

Db       88 CGKMYTRRGTFLEHAGDFDAGFFGIGPSEALAMDPQQRLLLEVSWEALERTGIDPTKLRG 147

Qy      125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184

Db      148 SATGVFAGVIHAGYGGQL-----SGELEGYGLTGSTLSVASGRVAYVLGLEGPAVSVDTA 202

Qy      185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244

Db      203 CSSSLVALHLAVQSLRSGECDLALAGGVTVMATPAAFVEFSRQRALARDGRCKVYAGAAD 262

Qy      245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304

Db      263 GTAWSEGAGVLVVERLVDARRLGHPVLALVRGSAVNQDGASNGLTAPNGPSQQRVIRAAL 322

Qy      305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364

Db      323 ASARLRAVEVDVVEGHGTGTMLGDPIEAQALLATY--GQDRVEPLWLGSIKSNIGHTSAA 380

Qy      365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424

Db      381 AGVAGVIKMVQAMRHGVMPKTLHVDVPTPHVDWSVGAVSLLTQPRAWSVHGRPRRAGVSS 440

Qy      425 FGMSGTNAHVVLEEAPAATCTPPAPERP----AELLVLSARTASALDAQAARLRDHLETY 480

Db      441 FGISGTNAHVILEQAPVVESVVPEVASPTAASAVPWVLSARSEQALAGQAQRLLAFVAAN 500

Qy      481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540

Db      501 PDLDPIDVGWSLVKTRAMFEHRAVVVGADRGALLAGLAALAAGESGAGVAVGRA-RSVGK 559

Qy      541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Db      560 TVFVFPGQGAQWVGMGAQLYAELPLFALAFDAVAEELDRHLRLPLRNVLW----EGDEAL 615

Qy      601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660

Db      616 LTSTEFAQPALFAIEVALATLLQHWGISPDFLIGHSVGEIAAAHLAGVLSLTDAAGLVAA 675

Qy      661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720

Db      676 RGRLMAELPAGGVMVVVAASEEEVLPVLVDGA---NLAAVNAPHSVVVSGCEAAVSDIAD 732

Qy      721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780

Db      733 HFARRGRRVHRLAVSHAFHSLLMEPMLAEFTRIAAGISVSKPRIPLVSNVTGQMAGAGYG 792

Qy      781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840

Db      793 DGQYWVEHARRPVRFAEGVQLLNAVGATRFVEVGPGGGLTALVEQSLPLGEALSVAMMRR 852

Qy      841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWID---------- 890

Db      853 EHPEVSSVLGAVATLFTAGAQMDWPAVFGSPGRRIELPTYAFQRQRYWLPPTSAGSADIS 912

Qy      891 ------------------------------------------------------------ 890

Db      913 GVGLLAARHGLLGAVVEQPDSDVVVLTGRLSVGEQRWLADHVIAGVVLLAGAAFVELALR 972

Qy      891 ------------------------------------------------------------ 890

Db      973 AADQVDCGVVEELTVVTPLVLPTVGGVQLQVVVGVGEMGQRPVSIYSRNAESDSGWVLHA 1032

Qy      891 ------------------------------------------------------------ 890

Db     1033 RGVLGAKAVAPAADLSVWPPLGAAPVDVDGAYQRFAELGYEYGRAFQGLTAMWRRESELF 1092

Qy      891 ------------------------------------------------------------ 890

Db     1093 ADVAVPDDVDVTLSGFGIHPLVLDAALHAMGMVGEQAATMLPFSWQGVSLHAAGASRVRA 1152

Qy      891 -----------------------------------------TKADDAARG--------- 899

Db     1153 RIAPAGDGTVSVELADQAGLPVLSVQALVMRSVSSQLLSAAVAAADAAGRGLLEVAWLPV 1212

Qy      900 -------------------DRRAP-----------------------------GAGHD 909

Db     1213 ELAHNDISADLVVWELESFQDGVGPVYSATHRVLVALQSWLAQERAGRLVLTQGSVGQD 1272

Qy      910 EVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGD---------------------- 945

Db     1273 ATNLAGAAVWGLVRSAQAEHP----GRVMLVDSDGSMDVGDVIGCGEEQLMIRNGTAYAA 1328

Qy      946 -----RP------------FRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFN 988

Db     1329 RLAQLRPQPILQLPDTNSGWRLVAGGAGALEDLTLASCPAKELAPGQVRIEVRALGVNFR 1388

Qy      989 DVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVT 1048

Db     1389 DVLVALGIYPGAAE--------LGAEGAGVVTEVGPGVTGLAVGDPVMGL-LGVAGSEAV 1439

Qy     1049 TSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQW 1108
```

2

```
Db       1440 VDARLVVKLPNRWPLTDAAGVPVVFLTAYYALRVLAQVQPGESVLVHAAAGGVGMAAVQL 1499

Qy       1109 AQHVGAEVHATAGTPEKRAYLESLGV--RYVSDSRSDRFVADVRAWTGGEGVDVVLNSLS 1166
              |:  | || |||  :   |      | ::|      :|:|||:   |        | | |||||||||:
Db       1500 ARLWGLEVFATA-SRGKWDTLHTMGCDNTHVADSRTLAFEETFWLTTEGRGVDVVLNSLA 1558

Qy       1167 GELIDKSFNLLRSHGRFVELGKRD 1190
              ||  |  |   ||    |||:|:|| :
Db       1559 GEFTDASLRLLPRGGRFIEMGKTE 1582
```

RESULT 5
AI2140
polyketide synthase [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AI2140
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2140
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2518 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB74379.1; PID:g17131773; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr2680

```
  Query Match             29.6%;  Score 2136;  DB 2;  Length 2518;
  Best Local Similarity   24.7%;  Pred. No. 3.5e-107;
  Matches  622;  Conservative  246;  Mismatches  515;  Indels 1132;  Gaps   38;

Qy          8 RAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGK 67
              :|  :|||||  || |||          | ||         |  :  :|::||     ::  :|  |||
Db         24 KAQTEPIAIVGMGCRFPGGADTPEKLWQLLRNGDDGITEIPSDRWSIETFYHNNPATPGK 83

Qy         68 TPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTET 127
                        |:  :    ||| ||||:|:||: :||  |||||| ||||||  | |   | |::|
Db         84 MSTRYGGFIGHLKEFDADFFGIAPKEAISLDPQQRLLLEVTWEALENTGILPEQLTGSQT 143

Qy        128 GVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSS 187
              ||||||· ::|    |        : :|||:   |    |   |||||| ||   || :||||| ||
Db        144 GVFIGISSNDYTTHLLNRPVT-DIDAYLATGNSHSTAAGRISYILGFTGPSLAVDTACSS 202

Qy        188 SLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFG 247
              ||||||||||||| |||    |: |||: :  ||     :   |: | || |||||  |    :||||
Db        203 SLVAVHLACQSLRQGECQLAVVGGVNRIFSPEFTINFSQARMLAADGRCKTFDTKADGFV 262

Qy        248 RGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADA 307
              ||||| |::|::|| |:|    :  |||||||||:|  || |||||||||  :|:  |::::|| :|
Db        263 RGEGCGVIILQKLSDAQAARNPILAVIRGSAVNQDGRSSGLTVPNGPAQQAVIRQALKNA 322

Qy        308 GCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGI 367
                   |  :  ||||||||| ||||||| || ||:|            :|||:||:|||:|| | |:||
Db        323 TVEPSDITYVEAHGTGTALGDPIEIGALGAVFGESHQ-TSPLLVGSLKTNIGHLEAAAGI 381

Qy        368 TGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGM 427
                ||:||||:|:||:|| ||| |    || | |          :    :      ||     |:  :  ||||||·
Db        382 AGLMKVVLALKHGEIPPHLHFQQPNPHIPWEQFPIQIPTTCTP---WHGKKLAGVSSFGF 438

Qy        428 SGTNAHVVLEEAPAATCTPP-APERPAELLVLSARTASALDAQAARLRDHLETYPSQCLG 486
              |||||||:|   :               :||  ::||  |||:| :||    |          :|: :     |
Db        439 SGTNAHVILAASEIQNQNKENTIKRPLQILNLSAKTETALKQLAQNYITYLKEHQDINLA 498

Qy        487 DVAFSLATTRSA--------MEHRLAVAA----TSREGLRAALDAAAQGQTSPGAVRSIA 534
              |: :|    | ||:                    |||.|:: |      |::|  |     |    ::      |       | |
Db        499 DICYSANTGRSSRGAASLTQFEHRLSIIANSITTAQEKLTCFINNEATTNLFTNKVNS-- 556

Qy        535 DSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLF--NQELDRPLREVMWAE 592
              || |:||||||||:|   ||   ||     |:|  |  |  |    |   ||  |  |:: :
Db        557 -PSRPKIAFLFTGQGSQYQNMGWELYQTEPFFKETIDKCCELLLANSGLD--LHSVLFPQ 613

Qy        593 PASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLE 652
                 :  || || :|||||| |||| || |||::|:   |||:|| ||||:|||||||
Db        614 DNENNHKLLSQTLYTQPALFILEYALCQLWLSWGIKPDFFMGHSVGEYVAACIAGVFSLE 673

Qy        653 DAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAG 712
              | : |:| | |||| ||   | ||:: |   :  :    :  |:   |||||:|||: ||:|
Db        674 DGLKLIATRARLMQQLPQ-GKMVAVAASPSQLQNFLLPYDQQVSIAAINAPNNTVISGEP 732

Qy        713 QPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSG 772
               :  | | : ::  :   | ||||||||:| ||  | :|  :::: |   ::||::|
Db        733 AAIEKIIAVLVSQNIQATPLSVSHAFHSPMMETMLGEFEKIAATINFHPPKYPIISNVTG 792

Qy        773 KACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARP 832
                   |:::| || || |: |:| ||: |            |:||| |  || :  : :  |
Db        793 SLINSEIATPEYWCRHIRQPVQFLAGVETLIQQNCSIFLEVGAKPILLQMARSIISQADK 852

Qy        833 AL-LASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPS-GGRRVPLPTYPWQRERYWID 890
               | | |  :  : :    :| ::  ::   |  : |      :  :|: || ||:|:|: :|||
Db        853 YLWLPSLQPKQQDWQVMLSSIAAIYGRGVKIDWHKFEQNYHHQRLDLPNYPFQRQNFWID 912

Qy        891 TK---------------------------------------------------------- 892
               |
Db        913 IKPVNKNYYSSSPNNIHPLLGQQLNLAKSSDIYFEQQLTNNNPDYLQDHKVFNQIILPGA 972

Qy        893 ------------------------------------------------------------ 892

Db        973 AYLEMSLAAAKTIFKNNSISVETVSFQETCILTPDNPKNIQFILKNNHEFEIVSTTSNHD 1032

Qy        893 ------------------------------------------------------------ 892

Db       1033 WITHATGKIKPHNNIPQPQQINLGELQNTFTKIIDITDFYQNLTQVGIEYGETFQAITHI 1092

Qy        893 --ADDAARGDRRAP---------------------------------------------- 904
                |: |  :  |
Db       1093 WYLDNQALAEIHLPSTCNCRDNYQFHPIILDACLQTIAAIFYSQPTPSNSSQDIENPSPN 1152

Qy        905 ------------------------------------------------------------ 904

Db       1153 LSPKRREALNVPPSLVGKGVRGLGLNTTNQVYLPIGVDKLTIYLPIGETVWSLVKLRTQH 1212

Qy        905 ------------------------------------------------------------ 904

Db       1213 KTAFALVADIQILSPSGQVIALVEGLQLKKIQSPNIKDWQNWLYEIEWRSQPLSITSDIS 1272

Qy        905 ------------------------------------------------------------ 904

Db       1273 LTSPAEISQELALKFTELLTTEEIKTYAQLLPQLEALSLYYIVQALQNLDVLSPEIAPQH 1332

Qy        905 ------------------------------------------------------------ 904

Db       1333 QRLYTHLLSLITNQREYEKPQLSPTQHPEIAAELTLIERCGTNLGKVLQGKCNPIDLLFP 1392

Qy        905 ---------------------------------------------GAG------------ 907
                                                           |||
Db       1393 NGDLTTLTQLYQNSPGAKVMNTLVQQAINSALKDLPPGEKVRILEIGAGTGGTTAYVLPQ 1452

Qy        908 --HDEVE----------------------------------------------------- 912
                 |  ||
Db       1453 LVHQSVEYVFTDISPLFLAKARQQFSEYEFVSYQTLNIEQPLTNQDITPHSFDIVIAANV 1512

Qy        913 ---------------------------EG-----------GAVRG--------------- 919
                                         ||           |   |
Db       1513 LHATENLNHTITNVKSLLNNQGLLIVLEGTIPSIWIDLIFGLTEGWWRFQDQDLRPHHPL 1572

Qy        920 ------------------------------------------------------------ 919

Db       1573 ISTTAWHSLLKTHDFTNIVNITPDSILPEALAQQSVIVAQNTPPQPSPYKREGEECLIIT 1632

Qy        920 ------------------------------------------------------------ 919
```

```
Db   1633 DLPAWGEAINQQLKLPSFTLKPSDNIAEFPHQSIKHIIYIAGCQDNITECNNILHLVQTL 1692

Qy    920 -------------------------------GDRRSARLDHP---------------- 930

Db   1693 IKTQHYPINLWLVTQGAISPHTITGLNQSSVWGMGKVIRLEHPELNCRCVDLDPTQELFT 1752

Qy    931 -----------PPESGR--------------REKVEAAGDRPFRLEIDEPGVLDHLVLRV 965

Db   1753 QVENLVTEITHPGEEAQITLHKSERKVARLTRCTLQILPHEPYRLTIAKRGVIASLKWQS 1812

Qy    966 TERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLGGECAGRIVAVGEG 1025

Db   1813 SPRRQPQPGEVEIQVKATGLNFIDVLDALGL----LPFERN---WFGVECAGEVVAIGEG 1865

Qy   1026 VNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIAR 1085

Db   1866 VTHLNVGDAVVALAADSFSQYVTTNANYVIKKPDSLSFTAAATIPANFLTAAYALREVAK 1925

Qy   1086 LQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRF 1145

Db   1926 IQPGKRILIHAATGGTGMAALQIAQQAGLEVFATASV-GKWETLRALGVQHIFNSRTLNF 1984

Qy   1146 VADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY-ADNQLGLRPFLR 1204

Db   1985 AEEIHEITQGEGVDIVFNSLSGEFIPASLSVLKPQGQFIEIGKRGVWNAQKVAQVKP--- 2041

Qy   1205 NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVF--------TPPPIATLPIARVADA 1256

Db   2042 DVVYHLVDLMSVAQQQPQTIQTLLHCLMAEFESGELRWSPTDRRSSHPQTIFPATKIVEA 2101

Qy   1257 FRSMAQAQHLGKLVLT-------------------------------------------- 1272

Db   2102 LQMMQQARHIGKIVITHTPTESLQPDATYLITGGMGGLGLRVARWLVEKGAKHLVLLGRN 2161

Qy   1273 ------------------------------------------------------------ 1272

Db   2162 QPDTTAQQQIQALEATGAKIIITQADVSQKDQLAAVLTDIQQNHPPLRGVIHAAGVLDDG 2221

Qy   1273 ---------------------------------------------LGDPE---------- 1277

Db   2222 ILQQLTPERSHRVMNPKVTGAWNLHQLTQDIPLDYFIMFSSAASLLGSPGQANHVAANTF 2281

Qy   1278 ----VQIR----IP----------------------THAGAGPSTGDR--DLLD---- 1299

Db   2282 LDALAQYRRHQGLPALSINWGVWSDIGAAAKRQVSNQMSSRGIGEITPDQGIDILEFILT 2341

Qy   1300 -------------------RLAS--------AAPAA--------------------RAAA 1312

Db   2342 QSSTQVGVIPINWSELLKQKLSSTFFADFYESTPSTSQEYQPSQLIQHIQCLGIKERISY 2401

Qy   1313 LEAFLRTQVSQVL-----RTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTF 1367

Db   2402 LRKYLQTEVSQVLGSSSSQLPNVQLG----FFDMGMDSLMMVELRSRLETSLKQTIPSTV 2457

Qy   1368 LSTSPNIALLAQNLLDALATALSLER--VAAENLRAGVQNDFVSSGADQDWEIIA 1420

Db   2458 LFEHPSI----QALAEYIATQLLPQETVISAINEQPDISSSDVSSKQEIETSITA 2508
```

RESULT 6
AB2012
hypothetical protein all1648 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AB2012
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AB2012
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1587 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB78014.1; PID:g17135468; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all1648

```
  Query Match             29.4%;  Score 2120;  DB 2;  Length 1587;
  Best Local Similarity   44.7%;  Pred. No. 1.4e-106;
  Matches  450;  Conservative  168;  Mismatches  337;  Indels   52;  Gaps   14;

Qy      6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65

Db     26 LERTQNEPIAIVGMGCRFPGDANNPEKFWELLRQGKDGITTVPPQRWDIDAYYDDDPDVP 85

Qy     66 GKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGT 125

Db     86 NKMYARYGGFINNVDQFDPQFFGITPREAIALDPQQRLLLEVSWEALENAGIAPQKLTGT 145

Qy    126 ETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAY 185

Db    146 QTGVFVGIGIDDY--AKRQIKHHIPIDAYTGSGNAFCFAAGRLSYLLGLQGPSLAIDTAC 203

Qy    186 SSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADG 245

Db    204 STSLVTIHLACQSLRNGECNLALAGGVSLMLSPEVTLYLSKTRALSPDGRCKTFDRDANG 263

Qy    246 FGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALA 305

Db    264 YVRGEGCGMVVLKRLSSAVADGDHILAVIRGSAVNQDGASSGLTVPNGTAQQAVIRQALA 323

Qy    306 DAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYAS 365

Db    324 NAKVTPAQISYLEAHGTGTALGDPIEVRAIDGVFGKGRSPNHPLILGSVKTNIGHLEIAA 383

Qy    366 GITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSF 425

Db    384 GMASLLKVILSLQHQEIPPHLNFQELNPDLAASAKSLKIPTSIIPWQPTEQPRMAGISSF 443

Qy    426 GMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQCL 485

Db    444 GLSGTNAHIIIEEPPQLTVTPAEVDRPLHVLMLSAKSEAALHTLATDWEHLLRNHPETNF 503

Qy    486 GDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSR-GKLAFL 544

Db    504 ADLAFSANTGRGSFNHRLAIIAQSTAQARQNLAAFNQKQPSLNVFSQEVEKGRQPKIAFL 563

Qy    545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604

Db    564 FTGQGSQYVGMGRQLYETQPTFRQALDECDRLLQPYLKESLLSVLYPQTPTAN-PLINQT 622

Qy    605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664

Db    623 AYTQTALFAIEYALCKLWQSWGIQPQGVLGHSVGEYVAACIAGVYTLQEGIELIAQRGQL 682

Qy    665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724

Db    683 MQALPQTGTMAAVFAPVETVARAIAPYANEVTIATINSPENVVISGVKAAIALVQADLTA 742

Qy    725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784

Db    743 QGIDVRPLQVSHAFHSPMMEPMLGEFKQVAAKINYQTPGIDWISSVTGAEITHSIDAE-Y 801

Qy    785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844

Db    802 WCQQIRDCVQFAPAMETLAQQGYDVLIEIGPHPVLTRLGKQTLSDPQILWLPSLHREQNN 861

Qy    845 PATVLEALGGLWAVGGLVSWAGLFPSG-GRRVPLPTYPWQRERYWIDTKADDAARGDRRA 903

Db    862 WQSLLQSVATLSVHGVRLDWSGFEQDYIRRRLTIPTYPFQRQRYWL-------AEAESTQ 914

Qy    904 PGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVL 963
```

```
Db        915 P-----EVIPVAAI-------------SPET---STIVAA------TETLESQILS-LVA 946

Qy        964 RVTERRAPGLGEVEIAVDA---AGLSFNDV---QLALGMVPDDLPGK 1004
              ::|      |:   ::::||    ||  : :    ||  |::     ||:
Db        947 KIT-----GMNPQQLSLDATLEGGLGLDSIMMTQLMNGIIKFIPPGQ 988


RESULT 7
S23070
erythronolide synthase (EC 2.3.1.94) II - Saccharopolyspora erythraea
N;Alternate names: 6-deoxyerythronolide B synthase II
C;Species: Saccharopolyspora erythraea
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000
C;Accession: S23070; S22011; S23205
R;Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
Eur. J. Biochem. 204, 39-49, 1992
A;Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of
A;Reference number: S23070; MUID:92155230; PMID:1740151
A;Accession: S23070
A;Molecule type: DNA
A;Residues: 1-3573 <BEV1>
A;Cross-references: EMBL:X62569
A;Experimental source: strain NRRL 2338
R;Bevitt, D.J.
submitted to the EMBL Data Library, September 1991
A;Reference number: S22011
A;Accession: S22011
A;Molecule type: DNA
A;Residues: 1-184,'I',186-301,'S',303-521,523-658,'A',660-993,1001-1212,'H',1214-1392,13
3479,'DH',3480-3572 <BEV2>
A;Cross-references: EMBL:X62569; NID:g46977; PIDN:CAA44448.1; PID:g581651
R;Caffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.
FEBS Lett. 304, 225-228, 1992
A;Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of th
A;Reference number: S23103; MUID:92316235; PMID:1618327
A;Accession: S23205
A;Molecule type: protein
A;Residues: 2-12,'XXX' <CAF>
A;Experimental source: strain CA340
C;Genetics:
A;Gene: eryA
A;Start codon: GTG
C;Function:
A;Description: catalyzes the construction of a polyketide chain, which is then cyclised
A;Pathway: erythromycin biosynthesis
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
ort-chain alcohol dehydrogenase homology
C;Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; multi
F;52-453/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;561-843/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1140-1308/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;1404-1475/Domain: acyl carrier protein homology <ACP1>
F;1519-1919/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2023-2305/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2857-3131/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;3149-3327/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;3420-3493/Domain: acyl carrier protein homology <ACP2>
F;1439/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

  Query Match             29.0%;  Score 2093.5;  DB 2;  Length 3573;
  Best Local Similarity   32.3%;  Pred. No. 1.1e-104;
  Matches  607;  Conservative  178;  Mismatches  532;  Indels  561;  Gaps   45;

Qy          2 ADRPIERAAE-DPIAIVGASCRLPGGVIDLSGFWTLL-EGSRDTVGRVPAER-WDAAAWF 58
              | |  :   | | :||||||  || |||:        | :| ||:  | |   ||:| ||       :
Db       1487 AVRVVGAADESEPIAIVGIGCRFPGGIGSPEQLWRVLAEGANLTTG-FPADRGWDIGRLY 1545

Qy         59 DPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIA 118
              |||| ||  :  |  :   ||:|  | |  ||||:||||| |||    ||:||   |||:| | |
Db       1546 HPDPDNPGTSYVDKGGFLTDAADFDPGFFGITPREALAMDPQQRLMLETAWEAVERAGID 1605

Qy        119 PSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPC 178
              | || ||:||||:|:       |    |    |    :    :| :  |||      || :|||:|   |    ||
Db       1606 PDALRGTDTGVFVGMNGQSYMQLL--AGEAERVDGYQGLGNSASVLSGRIAYTFGWEGPA 1663

Qy        179 VAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKA 238
              : ||||  |||||  :|||  |:||  ||||  ||||||::|    |   |   |       |    || ||||||
Db       1664 LTVDTACSSSLVGIHLAMQALRRGECSLALAGGVTVMSDPYTFVDFSTQRGLASDGRCKA 1723

Qy        239 FSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEI 298
              ||| ||||      ||  |  :||:  ||  |||:|  ::|||:||||:|  ||||:||    ||| |||
Db       1724 FSARADGFALSEGVAALVLEPLSRARANGHQVLAVLRGSAVNQDGASNGLAAPNGPSQER 1783

Qy        299 VLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNL 358
              |:::|||  :|    |:  |     |||||||  ||||||   ||  |   ||    ||       ||  :||||||:
Db       1784 VIRQALAASGVPAADVDVVEAHGTGTELGDPIEAGALIATYGQDRD--RPLRLGSVKTNI 1841

Qy        359 GHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPR 418
              ||  :   |:|   |::|||:::||  :|   |||    |:|  | |           :   |      |||     ||
Db       1842 GHTQAAAGAAGVIKVVLAMRHGMLPRSLHADELSPHIDWESGAVEVLREEVPWPAGERPR 1901

Qy        419 RAGVSSFGMSGTNAHVVLEEAP-----AATCTPPAPERPAELLVLSARTASALDAQAARL 473
              |||||||:|||||||::||||                 |  |       | |               |||  |:  :  :  |||    |
Db       1902 RAGVSSFGVSGTNAHVIVEEAPAEQEAARTERGPLP------FVLSGRSEAVVAAQARAL 1955

Qy        474 RDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSI 533
              :||       |       |  |  |::|||  |:    :  |  ||           |  |:  |  |||  |:|:    |    ||    :
Db       1956 AEHLRDTPELGLTDAAWTLATGRARFDVRAAVLGDDRAGVCAELDALAEGRPSADAVAPV 2015

Qy        534 ADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEP 593
                 :   |  |        :|   ||||  :||  |  |  |  :        |  |:          :          |      |  :|:   :
Db       2016 TSAPR-KPVLVFPGQGAQWVGMARDLLESSEVFAESMSRCAEALSPHTDWKLLDVVRGDG 2074

Qy        594 ASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLED 653
                      :|         ||  ||:     :||  |||:  ||  |     |  |||  ||:  ||  |||    |||
Db       2075 GPDPHERVD---VLQPVLFSIMVSLAELWRAHGVTPAAVVGHSQGEIAAAHVAGALSLEA 2131

Qy        654 AVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQ 713
                |     :||   |  ::::  |        |  |||:  |       ::        :|            |::||||    |       |:||
Db       2132 AAKVVALRSQVLRELDDQGGMVSVGASRDELETVLARWDGRVAVAAVNGPGTSVVAGPTA 2191

Qy        714 PVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGK 773
                :        |       ||    :  :  :   |  :|   |||   :|   :   :                :::     |  |:    |    |   ::|:
Db       2192 ELDEFFAEAEAREMKPRRIAVRYASHSPEVARIEDRLAAELGTITAVRGSVPLHSTVTGE 2251

Qy        774 ACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLL----------GLV 823
                          ||  |:   |       |          |       |  :       |       ||||  |      ||
Db       2252 VIDTSAMDASYWYRNLRRPVLFEQAVRGLVEQGFDTFVEVSPHPVLLMAVEETAEHAGAE 2311

Qy        824 PACMPDARPALLASSRAGRDEPATVLEAL--GGLWAVGGLVSWAGLFP--SGGRRVPLPT 879
              |:|                 :        |   :      |         |         |   :     ||    |          |  |     :|||   .  |||
Db       2312 VTCVP--------TLRREQSGPHEFLRNLLRAHVHGVG-----ADLRPAVAGGRPAELPT 2358

Qy        880 YPWQRERYW--------------------------------------------------- 888
              ||:: :|:|
Db       2359 YPFEHQRFWPRPHRPADVSALGVRGAEHPLLLAAVDVPGHGGAVFTGRLSTDEQPWLAEH 2418

Qy        889 ------------------------------------------------------------ 888

Db       2419 VVGGRTLVPGSVLVDLALAAGEDVGLPVLEELVLQRPLVLAGAGALLRMSVGAPDESGRR 2478

Qy        889 -IDTKADD-------------------AARGDR------------------------- 901
                ||  |:|                         |  |  |
Db       2479 TIDHAAEDVADLADAQWSQHATGTLAQGVAAGPRDTEQWPPEDAVRIPLDDHYDGLAEQG 2538

Qy        902 ----------RAPGAGHDEV----------EEG------------------GAV--------- 917
                                 ||          |         |||                          ||:
Db       2539 YEYGPSFQALRAAWRKDDSVYAEVSIAADEEGYAFHPVLLDAVAQTLSLGALGEPGGGKL 2598

Qy        918 -----------------------RGGDRRSARL------------------------DHPP 931
                                               | |   :  |:                                        :  |
Db       2599 PFAWNTVTLHASGATSVRVVATPAGADAMALRVTDPAGHLVATVDSLVVRSTGEKWEQPE 2658

Qy        932 PESGRRE-----------------------------------------KVEAAGDRP-- 947
              |   |   |                                           :  |   | | |
Db       2659 PRGGEGELHALDWVRLAEPGSTGRVVAADASDLDAVLRSGEPEPDAVLVRYEPEGDDPRA 2718
```

```
Qy      948 ------------------------------------------------------------ 947

Db     2719 AARHGVLWAAALVRRWLEQEELPGATLVIATSGAVTVSDDDSVPEPGAAAMWGVIRCAQA 2778

Qy      948 ------FRLEID-EPGVL--------------DHLVLRV-------------TERRAPG- 972
                  |: | |||:|              | | |:             |:| ||
Db     2779 ESPDRFVLLDTDAEPGMLPAVPDNPQLALRGDDVFVPRLSPLAPSALTLEAGTQRDVPGD 2838

Qy      973 -------------------LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012
                               ||| : | | |::| || ||||| |            :|
Db     2839 GAIDSVAFEPAPDVEQPLRRAGEVRVDVRATGVNFRDVLLALGMYPQKAD--------MG 2890

Qy     1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072
              | || : ||| |:    | |:  | ||||   |   |:     |   :|||:|:|
Db     2891 TEAAGVVTAVGPDVDAF-PGDRVLGLFQGAFAPIAVTDHRLLARVPDGWSDADAAAVPIA 2949

Qy     1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132
             | || |||  :| |: |: ||||| ||||:||| |:  ||| |||| |    | :|
Db     2950 YTTAHYALHDLAGLRAGQSVLIHAAAGGVGMAAVALARRAGAEVLATAG-PAKHGTLRAL 3008

Qy     1133 GV--RYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRD 1190
            |:   ::: ||   |     |  ||| |||||||||:|||:|:| :||   | |||:|| |
Db     3009 GLDDEHIASSRETGFARKFRERTGGRGVDVVLNSLTGELLDESADLLAEDGVFVEMGKTD 3068

Qy     1191 CYADNQLGLRPFLRNLSFSLVDLRGMMLERPA----------RVRALLEELLGLIAAGVF 1240
                        ||:      | ||    | |          |:  :| |::||: ||
Db     3069 ------------LRDAG----DFRG----RYAPFDLGEAGDDRLGEILREVVGLLGAGEL 3108

Qy     1241 TPPPIATLPIARVADAFRSMAQAQHLGKLVLTLG---DPEVQIRIPTHAGAGPSTGDRDL 1297
             |::   :       | :  |:: :|:| ||||      ||:   :  |           ||
Db     3109 DRLPVSAWELGSAPAALQHMSRGRHVGMLVLTQPAPVDPDGTVLI---------TGGTGT 3159

Qy     1298 LDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEA 1357
             | ||      ||      | :|                | :| | |: :| ||| ||
Db     3160 LGRL-----LARWLVTEHGVR----------------HLLVSRRGADAPGSDELRAEIE- 3197

Qy     1358 SLKLKLSTTFLSTSPNIALLAQNLLDALATAL----------------SLERVAAEN-LR 1400
             |          |:  || |  ||| |    |                |::  |  ||
Db     3198 DLGASAEIAACDTADRDALSA--LLDGLPRPLTGVVHAAGVLADGLVTSIDEPAVEQVLR 3255

Qy     1401 AGVQNDFVSSGADQDWEI 1418
             | |           |  | :
Db     3256 AKV---------DAAWNL 3264
```

RESULT 8
T17467
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17467
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3413 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology
C;Keywords: carrier protein
F;1608-1679/Domain: acyl carrier protein homology <ACP1>
F;3334-3405/Domain: acyl carrier protein homology <ACP2>

```
  Query Match             28.6%;  Score 2063.5;  DB 2;  Length 3413;
  Best Local Similarity   35.1%;  Pred. No. 4.3e-103;
  Matches  601;  Conservative  195;  Mismatches  547;  Indels  369;  Gaps   56;

Qy        8 RAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPG 66
              | |::|:|:|| :|| ||||       ||  |:  |  |  :   | :| |:    |||||| ||
Db       27 RHADEPLAVVGMACRFPGGVSSPEDLWQLVAGGVDALSDFPDDRGWELDGLFDPDPDHPG 86

Qy       67 KTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTE 126
              : ::  ||        |||  ||||||||| |||     |:|||  ||||:| : | :| |::
Db       87 TSYTSQGGFLRGAGLFDAGLFGISPREALVMDPQQRVLLETSWEALEDAGVDPLSLKGSD 146

Qy      127 TGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYS 186
             ||| |:    | |          : :::|  |:|   || :||:||  || || | :||| |
Db      147 VGVFSGVFTQGYGA----GAITPDLEAFAGIGAASSVASGRVSYVFGLEGPAVTIDTACS 202

Qy      187 SSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGF 246
            |||||:||| |:|:|:|||| |||||  ::| :| | |   |:  | || ||| ||||: |||
Db      203 SSLVAIHLAAQALRAGECSMALAGGATVMPTPGTFVAFSRQRVLAADGRSKAFSSTADGT 262

Qy      247 GRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALAD 306
            |  | |:|:||  |:       |||||:||||:| |||||:|||  ||| ||  ||: |:::||
Db      263 GWAEGAGVLVLERLSVAQERGHRILAVLRGSAVNQDGASNGLTAPNGPSQQRVIRKALAG 322

Qy      307 AGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASG 366
            ||  ||  |          |||||||| |||||  ||| || ||:          || :|||||:  || |:|
Db      323 AGLVASDVDVVEAHGTGTALGDPIEAQALLATYGQGRE--RPLWLGSVKSNFGHTQAAAG 380

Qy      367 ITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFG 426
             : |::|:| :|:|| :| ||        | :  |       : :          || :  |||||||:||
Db      381 VAGVIKMVQALRHGAMPPTLHVAEPTPEVDWSAGAVELLTEPREWPAGDRPRRAGVSAFG 440

Qy      427 MSGTNAHVVLEEAPAATCTPPAPERPAEL-LVLSARTASALDAQAARLRDHLETYPSQCL 485
             :|||||:::||||| |           ||       : ||:|| : ::| ||| || :  | :
Db      441 ISGTNAHLILEEAPPADAVAEEPEFKGPVPLVVSAGSPTSLAAQAGRLAEVLAS------ 494

Qy      486 GDV-----AFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
            | |       | |: |: :  |  | |  | :    |·| | |:|  :|| :        ||
Db      495 GGVSRARLASGLLSGRALLGDRAVVVAGTDEDAVAGLRALARGDRAPGVLTGSA--KHGK 552

Qy      541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFD-LCVRLFNQELD--------RPLREVMWA 591
               ||  |||:|  |||||  |||  |||   :                 ::|| 
Db      553 VVYVFPGQGSQRLGMGRELYDRYPVFATAFDEAC-----EQLDVCLAGRAGHRVRDVVLG 607

Qy      592 E-PASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFS 650
             | ||   :    ||:|| |||  ||  | || |  |   ||||  |::|  ||||||: ||  |||||
Db      608 EVPA--ETGLLNQTVFTQAGLFAVESALFRLAESWGVRPDVVLGHSIGEITAAYAAGVFS 665

Qy      651 LEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAG 710
             | ||   :||||||||||  |||||::  | ||:||  :         |  :||||  ||::|
Db      666 LPDAARIVAARGRLMQALAPGGAMVAVAASEAEVAELLGD---GVELAAVNGPSAVVLSG 722

Qy      711 AGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNL 770
                 |  | ||  |     ||  :|| |  ||||||| |||||  |         |::|  | |  :|||:
Db      723 DADAVVAAAARMRERGHKTKQLKVSHAFHSARMAPMLAEFAAELAGVTWREPEIPVVSNV 782

Qy      771 SGK-ACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPD 829
             :|:    |:  ||||   |    |||:|||| ||   :  |||:|| :  |  |
Db      783 TGRFAEPGELTEPGYWAEHVRRPVRFAEGVAAATESGGSLFVELGPGAALTALVE---ET 839

Qy      830 ARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI 889
             |    :|:  |  ||    :: |:   |:    |   |   |   |      ||| | : :: ||:
Db      840 AEVTCVAALRDDRPEVTALITAVAELFVRGVAVDWPALLPPVTGFVDLPKYAFDQQHYWL 899

Qy      890 D--TKADDAA-----RGDRRAPGA-------------------GHDEVEE---GG----- 915
              :| |||         |     ||                     |  :  :   ||
Db      900 QPAAQATDAASLGQVAADHPLLGAVVRLPQSDGLVFTSRLSLKSHPWLADHVIGGVVLVA 959

Qy      916 -------AVRGGDRRSARL---------------DH----------PPPESGRR-----EKV 940
                   ||| ||          :                   ||          | |:| |
Db      960 GTGLVELAVRAGDEAGCPVLEELVIEAPLVVPDHGGVRIQVVVGAPGETGSRAVEVYSLR 1019

Qy      941 EAAG------------------DRPF-----------RLEIDE--PGVLDH--------L 961
             | ||                       :||                |:::::    |::|
Db     1020 EDAGAEVWARHATGFLAATPSQHKPFDFTAWPPPGVERVDVEDFYDGLVDRGYAYGPSFR 1079

Qy      962 VLRVTERRAPGL-GEVEIA----VDAAGLSFN----DVQLALGM---VPDDLPGKPNPP- 1008
             ||   ||    :   || :|      |||       :     |   ||          : ||:|   |
Db     1080 GLRAVWRRGDEVFAEVALAEDDRADAARFGIHPGLLDAALHAGMAGATTTEEPGRPVLPF 1139
```

```
Qy       1009 ------LLLGGECAGRI-----------VAVGEGVNGLV------VGQPVIALSAGAFAT 1045
Db       1140 AWNGLVLHAAGASALRVRLAPSGPDALSVEAADEAGGLVVTADSLVSRPVSAEQLGAAAN 1199

Qy       1046 H----------VTTSAALVLPRPQALSAI-----EAAAMPVAYLTAWYALDRI--ARLQP 1088
Db       1200 HDALFRVEWTEISSAGDVPADHVEVLEAVGEDPLELTGRVLEAVQTWLA-DAADDARLVV 1258

Qy       1089 GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGV----RYVSDSRSDR 1144
Db       1259 VTRGAVHEVTDPAGAAV--W----GLIRAAQAENPDRIVLLDTDGEVPLGRVLATGEPQT 1312

Qy       1145 FV---------------ADVRAWTGGEGVDVVLNSLSGEL-IDKSFNLLRSHG--RFVEL 1186
Db       1313 AVRGATLFAPRLARAEAAEAPAVTGG----TVLISGAGSLGALTARHLVARHGVRRLVLV 1368

Qy       1187 GKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEE--------LLGLIAAG 1238
Db       1369 SRRGPDADGMAELTAELIAQGAEVAVVACDLADRD-QVRVLLAEHRPNAVVHTAGVLDDG 1427

Qy       1239 VFTPPPIATLPIARVADAFR-SMAQAQHLGKL---------------------------- 1269
Db       1428 VF-----ESLTRERLAKVFAPKVTAANHLDELTRELDLRAFVVFSSASGVFGSAGQGNYA 1482

Qy       1270 ----------------------------------VLTLGDPE------------------ 1277
Db       1483 AANAYLDAVVANRRAAGLPGTSLAWGLWEQTDGMTAHLGDADQARASRGGVLAISPAEGM 1542

Qy       1278 -----------VQIRI---PTHAG------------------AGP-STGDRDLLDRLASAA 1305
Db       1543 ELFDAAPDGLVVPVKLDLRKTRAGGTVPHLLRGLVRPGRQQARPASTVDNGLAGRLAGLA 1602

Qy       1306 PAARAAALEAFLRTQVSQVL--RTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKL 1363
Db       1603 PAEQEALLLDVVRTQVALVLGHAGPE-AVRADTAFKDTGFDSLTSVELRNRLREASGLKL 1661

Qy       1364 STTFLSTSPNIALLAQNLLDALATALSLERVA 1395
Db       1662 PATLVFDYPTPVALARYLRDELGDTVATTPVA 1693
```

RESULT 9
T17410
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17410
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3739 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800835; PIDN:AAC69330.1
C;Genetics:
A;Gene: pikAII
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1445-1516/Domain: acyl carrier protein homology <ACP1>
F;3570-3641/Domain: acyl carrier protein homology <ACP2>

```
  Query Match             28.5%;  Score 2052.5;  DB 2;  Length 3739;
  Best Local Similarity   28.0%;  Pred. No. 1.9e-102;
  Matches  604;  Conservative  190;  Mismatches  551;  Indels  811;  Gaps   36;

Qy          5 PIERAAED-PIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDP 62
Db       1532 PVDGGVDDEPVAIVGMACRLPGGVASPEDLWRLVAGGEDAISGFPQDRGWDVEGLYDPDP 1591

Qy         63 DAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSAL 122
Db       1592 DASGRTYCRAGGFLDEAGEFDADFFGISPREALAMDPQQRLLLETSWEAVEDAGIDPTSL 1651

Qy        123 VGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVD 182
Db       1652 QGQQVGVFAGTNGPHYEPLL--RNTAEDLEGYVGTGNAASIMSGRVSYTLGLEGPAVTVD 1709

Qy        183 TAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAE 242
Db       1710 TACSSSLVALHLAVQALRKGECGLALAGGVTVMSTPTTFVEFSRQRGLAEDGRSKAFAAS 1769

Qy        243 ADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKR 302
Db       1770 ADGFGPAEGVGMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRR 1829

Qy        303 ALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPE 362
Db       1830 ALADARLTTADVDVVEAHGTGTRLGDPIEAQALIATYGQGRDTEQPLRLGSLKSNIGHTQ 1889

Qy        363 YASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWP--DWNTPRRA 420
Db       1890 AAAGVSGIIKMVQAMRHGVLPKTLHVDRPSDQIDWSAGTVELLTEAMDWPRKQEGGLRRA 1949

Qy        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLV---LSARTASALDAQAARLRDHL 477
Db       1950 AVSSFGISGTNAHIVLEEAPVDE-DAPADEPSVGGVVPWLVSAKTPAALDAQIGR----L 2004

Qy        478 ETYPSQCL------GDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVR 531
Db       2005 AAFASQGRTDAADPGAVARVLAGGRAQFEHRAVALGTGQDDLAAALAA------PEGLVR 2058

Qy        532 SIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWA 591
Db       2059 GVA-SGVGRVAFVFPGQGTQWAGMGAELLDVSKEFAAAMAECEAALAPYVDWSLEAVVRQ 2117

Qy        592 EPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSL 651
Db       2118 APG---APTLERVDVVQPVTFAVMVSLAKVWQHHGVTPQAVVGHSQGEIAAAYVAGALSL 2174

Qy        652 EDAVFLVAARGRLMQALPAG-GAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAG 710
Db       2175 DDAARVVTLRSKSIGAHLAGQGGMLSLALSEAAVVERLAGF-DGLSVAAVNGPTATVVSG 2233

Qy        711 AGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNL 770
Db       2234 DPTQIQELAQACEADGVRARIIPVDYASHSAHVETIESELADVLAGLSPQTPQVPFFSTL 2293

Qy        771 SGKACTDEVSSPGYWVRHAREVVRFADGVKALHA-AGAGTFVEVGPKSTLLGLVPACMPD 829
Db       2294 EGAWITEPALDGGYWYRNLRHRVGFAPAVETLATDEGFTHFVEVSAHPVLTMALPETVTG 2353

Qy        830 ARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYW- 888
Db       2354 -----LGTLRRDNGGQHRLTTSLAEAWANGLTVDWASLLPTTTTHPDLPTYAFQTERYWP 2408

Qy        889 ------------------------------------------------------------ 888
Db       2409 QPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAVAGTVLL 2468

Qy        889 -----------------------------------------IDTKADD------------ 895
Db       2469 PGTAFVELAFRAGDQVGCDLVEELTLDAPLVLPRRGAVRVQLSVGASDESGRRTFGLYAH 2528

Qy        896 ------------------AARGDRRAP---------------------------GAGH-- 908
Db       2529 PEDAPGEAEWTRHATGVLAARADRTAPVADPEAWPPPGAEPVDVDGLYERFAANGYGYGP 2588

Qy        909 ------------DEV---------EEGGA------------------------VRGGDRRS 924
Db       2589 LFQGVRGVWRRGDEVFADVALPAEVAGAEGARFGLHPALLDAAVQAAGAGRGVRRGHAAA 2648

Qy        925 ARLDH------------------------------------------------------- 929
Db       2649 VRLERDLLYAVGATALRVRLAPAGPDTVSVSAADSSGQPVFAADSLTVLPVDPAQLAAFS 2708
```

```
Qy        930 ------------------------------------------------------------ 929
Db       2709 DPTLDALHLLEWTAWDGAAQALPGAVVLGGDADGLAAALRAGGTEVLSFPDLTDLVEAVD 2768

Qy        930 ----PPPES--------------------------------------------------- 934
Db       2769 RGETPAPATVLVACPAAGPDGPEHVREALHGSLALMQAWLADERFTDGRLVLVTRDAVAA 2828

Qy        935 -----------------GRREKVEA---------------AGDRP--------------- 947
Db       2829 RSGDGLRSTGQAAVWGLGRSAQTESPGRFVLLDLAGEARTAGDATAGDGLTTGDATVGGT 2888

Qy        948 ------------------------------------------------------------ 947
Db       2889 SGDAALGSALATALGSGEPQLALRDGALLVPRLARAAAPAAADGLAAADGLAALPLPAAP 2948

Qy        948 --FRLEIDEPGVLDHLVLRV--TERRAP---GLGEVEIAVDAAGLSFNDVQLALGMVPDD 1000
Db       2949 ALWRLEPGTDGSLESLTAAPGDAETLAPEPLGPGQVRIAIRATGLNFRDVLIALGMYPD- 3007

Qy       1001 LPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQA 1060
Db       3008 -------PALMGTEGAGVVTATGPGVTHLAPGDRVMGLLSGAYAPVVVADARTVARMPEG 3060

Qy       1061 LSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATA 1120
Db       3061 WTFAQGASVPVVFLTAVYALRDLADVKPGERLLVHSAAGGVGMAAVQLARHWGVEVHGTA 3120

Qy       1121 GTPEKRAYLESLGV--RYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLR 1178
Db       3121 SHGKWDA-LRALGLDDAHIASSRTLDFESAFRAASGGAGMDVVLNSLAREFVDASLRLLG 3179

Qy       1179 SHGRFVELGKRDCY-ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAA 1237
Db       3180 PGGRFVEMGKTDVRDAERVAADHPGVGYRAFDLGE------AGPERIGEMLAEVIALFED 3233

Qy       1238 GVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLG---DPEVQIRIPTHAGA------ 1288
Db       3234 GVLRHLPVTTWDVRRARDAFRHVSQARHTGKVVLTMPSGLDPEGTVLLTGGTGALGGIVA 3293

Qy       1289 ------------------------------------------------------------ 1288
Db       3294 RHVVGEWGVRRLLLVSRRGTDAPGAGELVHELEALGADVSVAACDVADREALTAVLDSIP 3353

Qy       1289 ------------------------------------------------------------ 1288
Db       3354 AEHPLTAVVHTAGVLSDGTLPSMTAEDVEHVLRPKVDAAFLLDELTSTPGYDLAAFVMFS 3413

Qy       1289 ------------------------------------------------------------ 1288
Db       3414 SAAAVFGGAGQGAYAAANATLDALAWRRRTAGLPALSLGWGLWAETSGMTGGLSDTDRSR 3473

Qy       1289 ------------------------------------------------------------ 1288
Db       3474 LARSGATPMDSELTLSLLDAAMRRDDPALVPIALDVAALRAQQRDGMLAPLLSGLTRGSR 3533

Qy       1289 -------------GPSTGDRDLLDRLASAAPAARAAALEAFLRTQVSQVL--RTPEIKV 1332
Db       3534 VGGAPVNQRRAAAGGAGEADTDLGGRLAAMTPDDRVAHLRDLVRTHVATVLGHGTPS-RV 3592

Qy       1333 GAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATA 1388
Db       3593 DLERAFRDTGFDSLTAVELRNRLNAATGLRLPATLVFDHPTPGELAGHLLDELATA 3648
```

RESULT 10
T17409
polyketide synthase type I - Streptomyces venezuelae
C;Species: Streptomyces venezuelae
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17409
R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A;Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A;Reference number: Z18773; MUID:98445333; PMID:9770448
A;Accession: T17409
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4613 <XUE>
A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800834; PIDN:AAC69329.1
C;Genetics:
A;Gene: pikAI
C;Superfamily: acyl carrier protein homology
C;Keywords: antibiotic biosynthesis; carrier protein
F;1010-1081/Domain: acyl carrier protein homology <ACP1>
F;2495-2566/Domain: acyl carrier protein homology <ACP2>
F;4407-4478/Domain: acyl carrier protein homology <ACP3>

```
  Query Match             28.5%;  Score 2051.5;  DB 2;  Length 4613;
  Best Local Similarity   34.6%;  Pred. No. 2.8e-102;
  Matches  589;  Conservative  183;  Mismatches  515;  Indels  417;  Gaps   52;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGK 67
Db       2601 ADDDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGEGITPFPTDRGWDLDGLYDADPDALGR 2660

Qy         68 TPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTET 127
Db       2661 AYVREGGFLHDAAEFDAEFFGVSPREALAMDPQQRMLLTTSWEAFERAGIEPASLRGSST 2720

Qy        128 GVFIGIGPSEYEAALPQATASAEIDAHGGL--GTMPSVGAGRISYALGLRGPCVAVDTAY 185
Db       2721 GVFIGLSYQDYAARVPNAPRGVE----GYLLTGSTPSVASGRIAYTFGLEGPATTVDTAC 2776

Qy        186 SSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADG 245
Db       2777 SSSLTALHLAVRALRSGECTMALAGGVAMMATPHMFVEFSRQRALAPDGRSKAFSADADG 2836

Qy        246 FGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALA 305
Db       2837 FGAAEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQALA 2896

Qy        306 DAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYAS 365
Db       2897 DARLAPGDIDAVETHGTGTSLGDPIEAQGLQATYGKERPAERPLAIGSVKSNIGHTQAAA 2956

Qy        366 GITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSF 425
Db       2957 GAAGIIKMVLAMRHGTLPKTLHADEPSPHVDWANSGLALVTEPIDWPAGTGPRRAAVSSF 3016

Qy        426 GMSGTNAHVVLEEAP---------------------------------AATCTPPAP-E 450
Db       3017 GISGTNAHVVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGTSEVAEGSEASEAPAAPGS 3076

Qy        451 RPAEL-----LVLSARTASALDAQAARLRDHLETYPSQC--------LGDVAFSLATTRS 497
Db       3077 REASLPGHLPWVLSAKDEQSLRGQAAALHAWLSEPAADLSDADGPARLRDVGYTLATSRT 3136

Qy        498 AMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSR-GKLAFLFTGQGAQTLGMG 556
Db       3137 AFAHRAAVTAADRDGFLDGLATLAQGGTS---AHVHLDTARDGTTAFLFTGQGSQRPGAG 3193

Qy        557 RGLYDVWSAFREAFD-LCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFE 615
Db       3194 RELYDRHPVFARALDEICAHL-DGHLELPLLDVMFAAEGSAEAALLDETRYTQCALFALE 3252

Qy        616 YALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMV 675
Db       3253 VALFRLVESWGMRPAALLGHSVGEIAAAHVAGVFSLADAARLVAARGRLMQELPAGGAML 3312

Qy        676 SIEAPEADVAAAVAP---HAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKAL 732
Db       3313 AVQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDADAAREAEAYWSGLGRRTRAL 3372

Qy        733 HVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAC-TDEVSSPGYWVRHARE 791
```

```
Db    3373 RVSHAFHSAHMDGMLDGFRAVLETVEFRRPSLTVVSNVTGLAAGPDDLCDPEYWVRHVRG 3432

Qy     792 VVRFADGVKALHAAGAGTFVEVGPKSTLLGLV-------------------PACMP---- 828

Db    3433 TVRFLDGVRVLRDLGVRTCLELGPDGVLTAMAADGLADTPADSAAGSPVGSPAGSPADSA 3492

Qy     829 ----DARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFP-SGGRRVPLPTYPWQ 883

Db    3493 AGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAGSGAHRVDLPTYSFR 3552

Qy     884 RERYWIDTKADDAARGDRRAPGAGHDEVEEGGAVRGGDRRSARLDHP-------PPESGR 936

Db    3553 RDRYWLDAPAADTA-------------VDTAGLGLG------TADHPLLGAVVSLPD--- 3590

Qy     937 REKVEAAGDRPFRLEI-DEPGVLDHLVLRVTERRAPGLGEVEIAVDAA-GLSFNDVQLAL 994

Db    3591 RDGLLLTG----RLSLRTHPWLADHAVLGSV--LLPGAAMVELAAHAAESAGLRDVRELT 3644

Qy     995 GMVPDDLP---------------GKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALS 1039

Db    3645 LLEPLVLPEHGGVELRVTVGAPAGEP------GGESA------GDGARPVSLHSRLADAP 3692

Qy    1040 AG-AFATHVTTSAALVLPRPQALSAIEAAAM---------------------PVAYLTAW 1077

Db    3693 AGTAWSCHAT--GLLATDRPELPVAPDRAAMWPPQGAEEVPLDGLYERLDGNGLAFGPLF 3750

Qy    1078 YALDRIARLQPGE---RVLIHAAT---------GGVGLAAVQWAQH-------------- 1111

Db    3751 QGLNAVWRYE-GEVFADIALPATTNATAPATANGGGSAAAAPYGIHPALLDASLHAIAVG 3809

Qy    1112 ---------------VGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFVADVRAWTGGE 1156

Db    3810 GLVDEPELVRVPFHWSGVTVHA-AGAAAARVRLASAGTDAVSLSLTD-----------GE 3857

Qy    1157 GVDVV-------------------LNSLSGELIDKSFNL---------LRSHGRFVELGK 1188

Db    3858 GRPLVSVERLTLRPVTADQAAASRVGGLMHRVAWRPYALASSGEQDPHATSYGPTAVLGK 3917

Qy    1189 RDC-------YADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFT 1241

Db    3918 DELKVAAALESAGVEVGLYPDLAALSQD-------------------------VAAG--A 3950

Qy    1242 PPP---IATLP--------------IARVADAFRSMAQAQHLG---KLVLTLG---DPEV 1278

Db    3951 PAPRTVLAPLPAGPADGGAEGVRGTVARTLELLQAWLADEHLAGTRLLLVTRGAVRDPE- 4009

Qy    1279 QIRIPTHAGAGPSTGDRDL-------LDRLASAAPAARAAAL-----------------E 1314

Db    4010 --------GSGADDGGEDLSHAAAWGLVRTAQTENPGRFGLLDLADDASSYRTLPSVLSD 4061

Qy    1315 AFLRTQVSQVLR------------TPEIKVGAEA-------------------------- 1336

Db    4062 AGLRDEPQLALHDGTIRLARLASVRPETGTAAPALAPEGTVLLTGGTGGLGGLVARHVVG 4121

Qy    1337 --------LFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATA 1388

Db    4122 EWGVRRLLLVSRRGTDAPGADELVHELEA------------LGADVSVAACDVADREALT 4169

Qy    1389 LSLERVAAEN------LRAGVQND 1406

Db    4170 AVLDAIPAEHPLTAVVHTAGVLSD 4193
```

RESULT 11
H70621
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: H70621
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70621
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2126 <COL>
A;Cross-references: GB:Z85982; GB:AL123456; NID:g3261718; PIDN:CAB06632.1; PID:g1839013
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: pks7
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I h
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein
F;56-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;563-845/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;1694-1875/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;1976-2051/Domain: acyl carrier protein homology <ACP2>

```
  Query Match             28.1%;  Score 2025;  DB 2;  Length 2126;
  Best Local Similarity   28.9%;  Pred. No. 2.8e-101;
  Matches  605;  Conservative  184;  Mismatches  567;  Indels  734;  Gaps   38;

Qy       3 DRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPD 61

Db      26 NRDLLARTTEPVAVVGMGCRYPGGVDSPETLWELVAHGRDAVSEFPADRGWDVAGLFDPD 85

Qy      62 PDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSA 121

Db      86 PDAVGKSYTRCGGFLTDVAGFDAEFFGIAPSEALAMDPQQRLLLEVSWEALERAGIDPIT 145

Qy     122 LVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAV 181

Db     146 LRGSQTGVFAGVFHGSYGG---QGRVPGDLERYGLRGSTLSVASGRVAYVLGLQGPAVSV 202

Qy     182 DTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSA 241

Db     203 DTACSSSLVALHLAVQSLRLGECDLALVGGVTVMATPAMFIEFSRQRALSADGRCKAYAG 262

Qy     242 EADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLK 301

Db     263 AADGTAFAEGAGVLVLARLADARRLGHPVLALVRGSAVNQDGASNGLATPNGPAQQRVIT 322

Qy     302 RALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHP 361

Db     323 AALASARLGVADVDVVEGHGTGTTLGDPIEAQAILATYG-QRPADRPLWLGSIKSNIGHT 381

Qy     362 EYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAG 421

Db     382 SAAAGVAGVIKMVQAMRHGVLPKTLHVDVPTPHVDWSAGAVSLLTEPRPWHVPGRPRRAG 441

Qy     422 VSSFGMSGTNAHVVLEEAPAATCTPPA--PERPAELLVLSARTASALDAQAARLRDHLET 479

Db     442 VSSFGISGTNAHVILEEAPAVEPVGAAHGNDPVAVPWVLSARSAQALTNQARRLLAWVGA 501

Qy     480 YPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRG 539

Db     502 DENVRPLDVGWSLVNTRSLFDHRAVVVGADRTQLMEGLTGLAAGVPGADVVAGRAQTV-G 560

Qy     540 KLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAA 599

Db     561 KTAFVFPGQGAQWLGMGAQLCATAPVFAEHIHRCERALREHVEWSLLDVLRGAPG---AP 617

Qy     600 LLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVA 659

Db     618 GLDRVDVVQPALWAVMVSLAELWRSVGVVPDAVIGHSQGEIAAAYVAGALSLRDAAAVVA 677

Qy     660 ARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIA 719

Db     678 LRSRLLVRLGGAGGMVSLACGQPQAEKLASQWGDRLNIAAVNGVSSVVLAGETDAVTELM 737

Qy     720 AAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEV 779
```

```
Db     738 QRCEAEGIRARRIDVDYASHSAQVDAIREELIAALRGIEPRTSTVAFFSTVTGELMDTAG 797

Qy     780 SSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPD------ARPA 833

Db     798 VNAEYWYRSIRQPVQFERAVRNAFDGGYRVFVESSPHPVLIAGIEETLVDCDRGATGEPI 857

Qy     834 LLASSRAGRDEPATVLEALGGLWAVGGL-------VSWAGLFPS-GGRRVPLPTYPWQRE 885

Db     858 VIPT--LGRDD-----GGVGRFWLSAGQAHVAGVGVDWRAAFADLGGRRVELPTYAFARQ 910

Qy     886 RYWID------------------------------------------------------- 890

Db     911 RFWLDGLGAVGGDLGGVGLVGAEHGLLAAVVQRPDSGGVVLTGRISVVAAPWLADHAVGP 970

Qy     891 ------------TKADD------------------------------------------- 895

Db     971 VVLFPGTGFVELALRAGDEVGCSVLQELTLQAPLVLPADGVRVQVVVGGVEQSGTRNVWV 1030

Qy     896 ----------------------------------------------------AARG---- 899

Db    1031 YSAAGQADSSPGWTLHAQGVLGVGSVQPAAELSVWPPVGARAMDVADGYQVLAARGYGYG 1090

Qy     900 ------------------------------------------------------------ 899

Db    1091 PAFRGLQALWRRGAEVFADVTLPEGVPIRGFGIHPAVLDAALHAWGIVEGEQQTMLPFSW 1150

Qy     900 -------------DRRAP-GAGHDEVE--------------------------------- 912

Db    1151 QGVCLHASGAARVRVRLAPVGRGAVSVELADPQGLPVLSVRQLMVRPVSAAALSRSTAGD 1210

Qy     913 -------------EG---------------GAVRGGD----------------------- 921

Db    1211 RGLLEMIWTPVPLEGGDIGDDAVVWELPPHAGAQAGGDVLAAVYRGVHEVLEVLQSWLAS 1270

Qy     922 --------------------------------RRSARLDHP------------------- 930

Db    1271 DATGLGVVVTRGAVGPVDDDVTDLAGAAVWGLVRSAQAEHPGRVVLVDTDGSVAVEDAVG 1330

Qy     931 -------PPESGRREKVEAAGDRP-------------FRLEIDEPGVLDHLVLRVTERR 969

Db    1331 FGARSGEPQLVVRRGRVYAARLAPVAAGLTLPSASAGGWRLVAGGGGTLADVVVAPVAPV 1390

Qy     970 APGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGL 1029

Db    1391 ELATGQVRVAVGAVGVNFRDVLVALGMYPGG--GE------LGVDGAGVVVEVGPGVTGL 1442

Qy    1030 VVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPG 1089

Db    1443 AVGDRVMGL-LGLVGSEAVVDARLVTMVPAGWSLVEAAAVPVAFLTAFYGLSVLAEVAAG 1501

Qy    1090 ERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVR--YVSDSRSDRFVA 1147

Db    1502 QKVLVHAGTGGVGMAAVSLARYWGAEVFVTA-SRAKWDTLRAMGFDDIHISDSRSLEFEE 1560

Qy    1148 DVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRD-----CYADNQLGLRPF 1202

Db    1561 AFLRATEGSGVDVVLNSLAGEFTDASLRLLPSGGRFIELGKTDIRDGQTVAERHRG---- 1616

Qy    1203 LRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ 1262

Db    1617 VRYRAFDLVE------AGPDRIAAMLSEVVGLLAAGVLARLPVKTFDARCAPAAYRFVSQ 1670

Qy    1263 AQHLGKLVLTL------------------------------------------------- 1273

Db    1671 ARHIGKVVLTIPDGPGGQSGLAGGTVVVTGGTGMAGSAVATHLVRRHGVANLVLVSRSGE 1730

Qy    1274 --------------GDPEVQI--------------------RIP---------------- 1283

Db    1731 QADRAAEVAALLREGGAQVAVVSCDVADRDALAALLAGLDPRYPLKGVFHAAGVLDDAVI 1790

Qy    1284 ------------------------------------------------------------ 1283

Db    1791 TGLTPDRVDTVLRAKVDGAWNLHELTEDMDLSAFVVFSSMAGIVGTPAQGNYAAANAFLD 1850

Qy    1284 --------------------------------------THAGAGPSTGDRDL--LD--- 1299

Db    1851 GLVAYRRSRGLAGLSVAWGLWEQASAMTRHLGERDRARMTQAGLAPLTTEQALGFLDTAL 1910

Qy    1300 ----------RLASAAPAARAAALEAFLR---------------TQVSQ---VLR----T 1327

Db    1911 QADRAVVVAARLDRAALAGAGAALPALFSQLAAGPTRRRIDAADTAVSMSGLVSRLHALT 1970

Qy    1328 PE----------------------IKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLS 1364

Db    1971 PERRQRELTDLVISNAAAVLGRSSSVDINAHKAFQDLGFDSLTAVELRNRLKTATGLTLS 2030

Qy    1365 TTFLSTSPNIALLAQNLLDALATALSLERVAAENLRAGVQNDFVSSGADQ 1414

Db    2031 PTLIFDYPTPATLAEHLDSRLVTA--------------------SGSDQ 2059
```

RESULT 12
A70984
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jul-2000
C;Accession: A70984
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70984
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2188 <COL>
A;Cross-references: GB:Z83857; GB:AL123456; NID:g3242252; PIDN:CAB06099.1; PID:g3242253
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: ppsC
C;Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I hc
nase homology; [acyl-carrier-protein] S-malonyltransferase.homology
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;56-460/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;572-853/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1482-1770/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;1802-1982/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;2072-2141/Domain: acyl carrier protein homology <ACP1>
F;2105/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

```
  Query Match             27.9%;  Score 2014.5;  DB 2;  Length 2188;
  Best Local Similarity   31.7%;  Pred. No. 1.1e-100;
  Matches  565;  Conservative  172;  Mismatches  511;  Indels  537;  Gaps   35;

Qy       7 ERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPG 66

Db      30 EKSSSEPIAVIGMGCRFPGGVNNPEQFWDLLCAGRSGIVRVPAQRWDADAYYCDDHTVPG 89

Qy      67 KTPVTRASFLSDVAC--FDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124

Db      90 TICSTEGGFLTSWQPDEFDAEFFSISPREAAAMDPQQRLLIEVAWEALEDAGVPQHTIRG 149

Qy     125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184

Db     150 TQTSVFVGVTAYDYMLTLAGRLRPVDLDAYIPTGNSANFAAGRLAYILGARGPAVVIDTA 209

Qy     185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244

Db     210 CSSSLVAVHLACQSLRGRESDMALVGGTNLLLSPGPSIACSRWGMLSPEGRCKTFDASAD 269

Qy     245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304

Db     270 GYVRGEGAAVVVLKRLDDAVRDGNRILAVVRGSAVNQDGASSGVTVPNGPAQQALLAKAL 329
```

```
Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364
Db        330 TSSKLTAADIDYVEAHGTGTPLGDPIELDSLSKVFS-DRAGSDQLVIGSVKTNLGHLEAA 388

Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424
Db        389 AGVAGLMKAVLAVHNGYIPRHLNFHQLTPHASEAASRLRIAADGIDWPTTGRPRRAGVSS 448

Qy        425 FGMSGTNAHVVLEEAP---AATCTPP--APERPAELLVLSARTASALDAQAARLRDHLET 479
Db        449 FGVSGTNAHVVIEQAPDPMAAAGTEPQRGPVPAVSTLVVFGKTAPRVAATASVLADWLDG 508

Qy        480 YPSQC--LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSS 537
Db        509 -PGAAVPLADVAHTLNHHRARQTRFGTVAAVDRRQAVIGLRALAAGQSAPGVVAPREGSI 567

Qy        538 RGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVD 597
Db        568 GGGTVFVYSGRGSQWAGMGRQLLADEPAFAAAIAELEPEFVAQGGFSLRDVI---AGGKE 624

Qy        598 AALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFL 657
Db        625 LVGIEQ---IQLGLIGMQLALTALWRSYGVTPDAVIGHSMGEVAAAVVAGALTPAQGLRV 681

Qy        658 VAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAG------- 710
Db        682 TAVRSRLMAPLSGQGTMALLEL-DAEATEALIADYPEVSLGIYASPRQTVISGPPLLIDE 740

Qy        711 ---------------------------AGQP---------------------------- 714
Db        741 LIDKVRQQNGFATRVNIEVAPHNPAMDALQPAMRSELADLTPQPPTIPIISTTYADLGIS 800

Qy        715 ---------------------------------------------VHAIAAAMAA----- 724
Db        801 LGSGPRFDAEHWATNMRNPVRFHQAIAHAGADHHTFIEISAHPLLTHSISDTLRASYDVD 860

Qy        725 ---------RGAR--------------------------------TKALHVSH------- 736
Db        861 NYLSIGTLQRDAHDTLEFHTNLNTTHTTHPPQTPHPPEPHPVLPTTPWQHTQHWITATSA 920

Qy        737 AFHSPLMAPML---------------------------------------EAFGRVA---- 754
Db        921 AYHRPDTHPLLGVGVTDPTNGTRVWESELDPDLLWLADHVIDDLVVLPGAAYAEIALAAA 980

Qy        755 -----------------ESVSYRRPSIVLVSNLSG--KACTDEV---SSPGYWVRHA- 789
Db        981 TDTFAVEQDQPWMISELDLRQMLHVTPGTVLVTTLTGDEQRCQVEIRTRSGSSGWTTHAT 1040

Qy        790 ----------------REVVRFAD------------------------------------ 797
Db       1041 ATVARAEPLAPLDHEGQRREVTTADLEDQLDPDDLYQRLRGAGQQHGPAFQGIVGLAVTQ 1100

Qy        798 -GVK-------------------------ALHAAGA---------GTFVEVGPKSTLLG 821
Db       1101 AGVARAQVRLPASARTGSREFMLHPVMMDIALQTLGATRTATDLAGGQDARQGPSSNSAL 1160

Qy        822 LVPA-----------------------------------------CMPDARPALLASS 838
Db       1161 VVPVRFAGVHVYGDITRGVRAVGSLAAAGDRLVGEVVLTDANGQPLLVVDEVEMAVLGSG 1220

Qy        839 RAGRD----------EPATV---LEALGGLWAVG------------------------- 859
Db       1221 SGATELTNRLFMLEWEPAPLEKTAEATGALLLIGDPAAGDPLLPALQSSLRDRITDLELA 1280

Qy        860 -----------GLVSWAGL------------FPS-----------------------G 871
Db       1281 SAADEATLRAAISRTSWDGIVVVCPPRANDESMPDEAQLELARTRTLLVASVVETVTRMG 1340

Qy        872 GRRVPLPTYPWQRERYWIDTK-------------ADDAARGDRRAPGAGHDE-------V 911
Db       1341 ARKSP---------RLWIVTRGAAQFDAGESVTLAQTGLRGIARVLTFEHSELNTTLVDI 1391

Qy        912 EEGG-------------------AVRGGDRRSARLDHPPPESGRREKVEA--------- 942
Db       1392 EPDGTGSLAALAEELLAGSEADEVALRDGQRYVNRLVPAPTTTSGDLAAEARHQVVNLDS 1451

Qy        943 --AGDRPFRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDD 1000
Db       1452 SGASRAAVRLQIDQPGRLDALNVHEVKRGRPQGDQVEVRVVAAGLNFSDVLKAMGV---- 1507

Qy       1001 LPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQA 1060
Db       1508 YPGLDGAAPVIGGECVGYVTAIGDEVDGVEVGQRVIAFGPGTFGTHLGTIADLVVPIPDT 1567

Qy       1061 LSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATA 1120
Db       1568 LADNEAATFGVAYLTAWHSLCEVGRLSPGERVLIHSATGGVGMAAVSIAKMIGARIYTTA 1627

Qy       1121 GTPEKRAYLESLGVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSH 1180
Db       1628 GSDAKREMLSRLGVEYVGDSRSVDFADEILELTDGYGVDVVLNSLAGEAIQRGVQILAPG 1687

Qy       1181 GRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVF 1240
Db       1688 GRFIELGKKDVYADASLGLAALAKSASFSVVDLDLNLKLQPARYRQLLQHILQHVADGKL 1747

Qy       1241 TPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTH 1285
Db       1748 EVLPVTAFSLHDAADAFRLMASGKHTGKIV---------ISIPQH 1783
```

RESULT 13
T30228
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30228
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staur
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30228
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6260 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987102; PIDN:CAA60462.1
C;Genetics:
A;Gene: rapC
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;55-451/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1536-1607/Domain: acyl carrier protein homology <ACP1>
F;1651-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;3123-3194/Domain: acyl carrier protein homology <ACP2>
F;3238-3632/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3725-3999/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;5114-5185/Domain: acyl carrier protein homology <ACP3>
F;5242-5638/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;5759-6044/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;6135-6206/Domain: acyl carrier protein homology <ACP4>

```
  Query Match             27.9%;  Score 2013.5;  DB 2;  Length 6260;
  Best Local Similarity   31.0%;  Pred. No. 4.8e-100;
  Matches  569;  Conservative  192;  Mismatches  551;  Indels  523;  Gaps   38;

Qy         11 EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGKTP 69
Db       3216 DEPLAVVGMACRLPGGVSSPEDLWRLVESGTDAISGFPADRGWDAESLFDPDPDASGKSY 3275

Qy         70 VTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTET-- 127
Db       3276 CVEGGFLDSAGSFDAGFFGISPREALAMDPQQRLIMEVSWEAFERAGIEPGSVRGTHRRL 3335

Qy        128 --GVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAY 185
```

```
Db   3336 HGRVRGGYG-----------AGADLGGFAATASATSVLSGRVSYFFGLEGPAITVDTAC 3383

Qy    186 SSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADG 245
Db   3384 SSSLVALHQAGYALRQGECSLALVGGVTVMATPQSFVEFSRQRGLASDGRCKAFADSADG 3443

Qy    246 FGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALA 305
Db   3444 TGWAEGVGVLLVERLSDAQAKGHQVLAVVRSSAVNQDGASNGLTAPNGPSQQRVIQAALS 3503

Qy    306 DAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYAS 365
Db   3504 NAGLAAHEVDVVEAHGTGTTLGDPIEAQALIATYGQDRE--RPLLLGSLKSNIGHAQAAS 3561

Qy    366 GITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSF 425
Db   3562 GVSGVIKMVMALQHNTVPRTLHVDEPSRHVDWAAGAVELVRENQPWPGTDRPRRAGVSSF 3621

Qy    426 GMSGTNAHVVLEEAPAATCTPPAPE-RPAEL---------LVLSARTASALDAQAARLRD 475
Db   3622 GVSGTNAHVILESAPPA---QPAEEAQPVETPVVASDVLPLVISAKTQPALTEHEDRLRA 3678

Qy    476 HLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQ--TSPGAVRSI 533
Db   3679 YLAASPGADIRAVASTLAVTRSVFEHRAVLLG----------DDTVTGTAVTDP------ 3722

Qy    534 ADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEP 593
Db   3723 ------RIVFVFPGQGWQWLGMGSALRDSSVVFAERMAECAAALREFVDWDLFTVL---- 3772

Qy    594 ASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLED 653
Db   3773 --DDPAVVDRVDVVQPASWAMMVSLAAVWQAAGVRPDAVIGHSQGEIAAACVAGAVSLRD 3830

Qy    654 AVFLVAARGR-LMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAG 712
Db   3831 AARIVTLRSQAIARGLAGRGAMASVALPAQDVEL-----VDGAWIAAHNGPASTVIAGTP 3885

Qy    713 QPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSG 772
Db   3886 EAVDHVLTAHEAQGVRVRRITVDYASHTPHVELIRDELLDITSDSSSQTPLVPWLSTVDG 3945

Qy    773 KACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARP 832
Db   3946 -TWVDSPLDGEYWYRNLREPVGFHPAVSQLQAQGDTVFVEVSASPVLLQAM-----DDDV 3999

Qy    833 ALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRV-PLPTYPWQRERYWIDT 891
Db   4000 VTVATLRRDDGDATRMLTALAQAYVHGVTVDWPAILGTTTTRVLDLPTYAFQHQRYWLKS 4059

Qy    892 -------------------------------------KADDAARGDRRAPGAGH------ 908
Db   4060 VDRAAADGHPLLGTVVALPGSDGVVLTGRVSLATHTWLADHAVRGSVLLPGTAFVELVVR 4119

Qy    909 --DEVE------------------------------------------------------ 912
Db   4120 AADEVECDVVDELVIETPLLLPQTGGVQLSVSVAEADESGHRTVTVFSQADNTDAWIRHV 4179

Qy    913 ---------------------------------------EG-----------GAVRGGD-- 921
Db   4180 SATISTSDTPLSLPESDTWPPAQAQPMNVADFYDRLMAEGYEYGPAFQGLQAAWRDGDTV 4239

Qy    922 -----------RRSARLD-HPP---------------------PESGRREKVEAAGD-- 945
Db   4240 YAEVALAEEQAQEAARYDVHPALLDAALHACTLNTPDAEEGVGLPFSWNHVQVHATGPAT 4299

Qy    946 ------------------------------------------------------------ 945
Db   4300 LRVAVTQEADGWSLRVADDTGRPVATIGSLITRPVTADTLGSTTDDLLTLTWTVIPPPQE 4359

Qy    946 --------------------------------------------------------RPFR 949
Db   4360 TSLPMVRYEDLADGNMPVPDVVVFTALPNSSENPLGPRPAGPADPLAQPDTDRQVLKPFQ 4419

Qy    950 ----------------------------------------------------------LE 951
Db   4420 TWLPGTFTDSTLVVRTGTGLAAAAVSGLMRSAQSEHPGRFILVESDDDTLTLRQLAATVG 4479

Qy    952 IDEP--------------------------------------GVLDHLVLRVTERRA 970
Db   4480 LDEPRLRINDGRYEAARLARTNTTPEDTPLTIPDSRAWLLEQPRSGTLQDLALVPTDTAE 4539

Qy    971 PGL--GEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNG 1028
Db   4540 QPLRPGEVRIDVRAAGLNFRDVVVALGMVNDNRP--------TGGEAAGVVLEVGPDVQD 4591

Qy   1029 LVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP 1088
Db   4592 LAPGDRVFGVVGGSFGPAAIADRRMLAVIPDGWSFTTAASVPVVFATAYYGLVDLAGLSE 4651

Qy   1089 GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLES-LGVRYVSDSRSDRFVA 1147
Db   4652 GESVLIHAAAGGVGMAATQIARHLGARIYATASTGKQHILREAGLEDTHIADSRTLSFQE 4711

Qy   1148 DVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGL-RPFLRNL 1206
Db   4712 TFLNNTHGQGVDVVLNSLSGDFVDASLDLLPRGGRFVEMGKTDIRDPHQVTADRPGTTYQ 4771

Qy   1207 SFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHL 1266
Db   4772 AFDLMD------AGPDRLREIITELLTLFTQGVLLPLPVQAWDIRQARDAFSWMSRARHI 4825

Qy   1267 GKLVLTL---GDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQVSQ 1323
Db   4826 GKIVLTIPRRPDPDGTILIT--GGSGVLAG---ILARHLAAEHGARHLLL--LSRTTPDQ 4878

Qy   1324 VLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLD 1383
Db   4879 ALIKELAELGAH-------VDTATC-------DVSDRAGLARVLAGVSPEHPLTA----- 4919

Qy   1384 ALATALSLERVAAENLRAGVQNDFVSSGADQDWEI 1418
Db   4920 VIHTAGALDDGVVESLTTQQLDTVLRPKADGAWHL 4954
```

RESULT 14.
T17463
rifamycin polyketide synthase modules 1-3 - Amycolatopsis mediterranei
C;Species: Amycolatopsis mediterranei
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C;Accession: T17463
R;Schupp, T.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z18802
A;Accession: T17463
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4735 <SCH>
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227120; PIDN:CAA11035.1
A;Experimental source: strain LBG A3136
C;Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C;Keywords: carrier protein
F;53-500/Domain: acetate-CoA ligase homology <ACL>
F;543-610/Domain: acyl carrier protein homology <ACP1>
F;2102-2173/Domain: acyl carrier protein homology <ACP2>
F;3079-3150/Domain: acyl carrier protein homology <ACP3>
F;4578-4649/Domain: acyl carrier protein homology <ACP4>

```
  Query Match             27.8%;  Score 2002.5;  DB 2;  Length 4735;
  Best Local Similarity   37.9%;  Pred. No. 1.3e-99;
  Matches  586;  Conservative  183;  Mismatches  526;  Indels  251;  Gaps   50;

Qy      2 ADRPIE---RAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAW 57
Db   2182 ANRPAEIGTAAAEEPIAIVAMACRFPGGVHSPEDLWRLVADGADAVTEFPADRGWDTDRL 2241
```

```
Qy      58 FDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAI 117
Db    2242 YHEDPDHEGTTYVRHGAFLDDAAGFDAAFFGISPNEALAMDPQQRLLLETSWELFERAAI 2301

Qy     118 APSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGP 177
Db    2302 DPTTLAGQDIGVFAGVNSHDYSMRMHRA---AGVEGFRLTGGSASVLSGRVAYHFGVEGP 2358

Qy     178 CVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCK 237
Db    2359 AVTVDTACSSSLVALHMAVQALQRGECSMALAGGVMVMGTVETFVEFSRQRGLAPDGRCK 2418

Qy     238 AFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQE 297
Db    2419 AFADGADGTGWSEGVGLLLVERLSEAQRRGHQVLAVVRGSAVNSDGASNGLTAPNGPSQQ 2478

Qy     298 IVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTN 357
Db    2479 RVIRKALAAAGLSTSDVDAVEAHGTGTTLGDPIEAEALLATYGQNRE--TPLWLGSVKSN 2536

Qy     358 LGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISW--GDLRLTVTRARTPWPDWN 415
Db    2537 LGHTQAAAGVAGVIKMVMAMRHGVLPRTLHVDRPSSYVDWSAGAVEL-LTEAR-DWVSNG 2594

Qy     416 TPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLV---LSARTASALDAQAAR 472
Db    2595 HPRRAGVSSFGIGGTNAHVVLEEV-AAPITTPQPE-PAEFLVPVLVSARTAAGLRGQAGR 2652

Qy     473 LRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRS 532
Db    2653 LAAFLGDRTDVRVPDAAYALATTRAQLDHRAVVLASDRAQLCADLAAFGSGVVTGTPV-- 2710

Qy     533 IADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQEL-DRPLREVMWA 591
Db    2711 -----DGKLAVLFTGQGSQWAGMGRELAETFPVFRDAFEAACEAVDTHLRERPLREVVFD 2765

Qy     592 EPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSL 651
Db    2766 -----DSALLDQTMYTQGALFAVETALFRLFESWGVRPGLLAGHSIGELAAAHVSGVLDL 2820

Qy     652 EDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGA 711
Db    2821 ADAGELVAARGRLMQALPAGGAMVAVQATEDEVAPLL---DGTVCVAAVNGPDSVVLSGT 2877

Qy     712 GQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLS 771
Db    2878 EAAVLAVADELAGRGRKTRRLAVSHAFHSPLMEPMLDDFRAVAERLTYRAGSLPVVSTLT 2937

Qy     772 GKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDAR 831
Db    2938 GELAA--LDSPDYWVGQVRNAVRFSDAVTALGAQGASTFLELGPGGALAAMALGTLGGPE 2995

Qy     832 PALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLF--PSGGRRVPLPTYPWQRERYWI 889
Db    2996 QSCVATLRKNGAEVPDVLTALAELHVRGVGVDWTTVLDEPATAVGTVLPTYAFQHQRFWV 3055

Qy     890 DTKADDAARGDRRAPGAGH----------DEVEEGGAVRGG--DRRSARLDHPPPESG-- 935
Db    3056 D--VDETAAVSVTPPPAEPIVDRPVQDVLELVRESAAVVLGHRDAGSFDLDRSFKDHGFD 3113

Qy     936 -------RREKVEAAG-DRPFRLEIDEPG---VLDHLVLRVTERRAPGLGEVEIAVDAAG 984
Db    3114 SLSAVKLRNRLRDFTGVELPSTLIFDYPNPAVLADHL-------RAELLGERPAAPAPVT 3166

Qy     985 LSFNDVQLALGMVPDDLPGKPNPP-----LLLGGECA--GRIVAVGEGVNGLVVGQPVIA 1037
Db    3167 RDVSDEPIAIVGMSTRLPGGADSPEELWKLVAEGRDAVSGFPVDRGWDLDGLYHPDPAHA 3226

Qy    1038 LSAGAFATHVTTSAALVLPRPQ------ALSAIEAAAM-PVAYL---TAWYALDRIARLQ 1087
Db    3227 ------GTSYTRSGGFLHDAAQFDAGLFGISPREALAMDPQQRLLLETSWEALER-AGVD 3279

Qy    1088 P----GERV-----LIH---------------------------------------AAT 1098

Db    3280 PLSARGSDVGVFTGIVHHDYVTRLREVPEDVQGYTMTGTASSVASGRVAYVFGFEGPAVT 3339

Qy    1099 GGVGLAAVQWAQHVGAEV-------HATAG------TPEKRAYLESLGVRYVS-DSRSDR 1144
Db    3340 VDTACSSSLVAMHLAAQALRQGECSMALAGGATVMASPD--AFLEFSRQRGLSADGRCKA 3397

Qy    1145 FV--ADVRAWTGGEGVDV-------------VLNSLSGELIDK---SFNLLRSHGRFVEL 1186
Db    3398 YAEGADGTGWAEGVGVVVLERLSVARERGHRVLAVLRGSAVNQDGASNGLTAPNGPSQQR 3457

Qy    1187 GKRDCYADNQLGLRPFLRNLSFSLVDL-----RGMMLERPARVRALL-------EELLGL 1234
Db    3458 VIRGALA--SAGLAP-------SDVDVVEGHGTGTALGDPIEVQALLATYGQEREQPLWL 3508

Qy    1235 -----------IAAGVFTPPPIATLPIARVADAFRS--MAQAQHLGKLVLTLGDPEVQIR 1281
Db    3509 GSLKSNLGHTQAAAGV--------VGVIKMIMAMRHGVMPATLHVDERTSQVDWSAGAIE 3560

Qy    1282 IPTHAGAGPSTG-------------------------DRDLLDRLASAAP---AARAAA 1312
Db    3561 VLTEAREWPRTGRPRRAGVSSFGASGTNAHLIIEEGPAEEAVDEEVASVVPLVVSARSAG 3620

Qy    1313 LEAFLRTQVSQVLRTPEIKVGAEALFT-RLGMDSLMAVELRNRIEA 1357
Db    3621 SLAGQAGRLAAVLENESLAGVAGALVSGRATLNERAVVIAGSRDEA 3666
```

RESULT 15
T30225
polyketide synthase - Streptomyces hygroscopicus
C;Species: Streptomyces hygroscopicus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C;Accession: T30225
R;Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Stau
Gene 169, 9-16, 1996
A;Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy
A;Reference number: Z20782; MUID:96186896; PMID:8635756
A;Accession: T30225
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-10223 <APA>
A;Cross-references: EMBL:X86780; NID:g987088; PID:g987099; PIDN:CAA60459.1
C;Genetics:
A;Gene: rapB
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C;Keywords: carrier protein
F;54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;1533-1604/Domain: acyl carrier protein homology <ACP1>
F;1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;3172-3243/Domain: acyl carrier protein homology <ACP2>
F;3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F;3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;5160-5231/Domain: acyl carrier protein homology <ACP3>
F;5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F;6776-6847/Domain: acyl carrier protein homology <ACP4>
F;6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F;8411-8482/Domain: acyl carrier protein homology <ACP5>
F;8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F;9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F;10069-10140/Domain: acyl carrier protein homology <ACP>

```
  Query Match             27.7%;  Score 2000.5;  DB 2;  Length 10223;
  Best Local Similarity   30.5%;  Pred. No. 4.6e-99;
  Matches  562;  Conservative  197;  Mismatches  562;  Indels  521;  Gaps   35;

Qy       5 PIERAAED-PIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDP 62
Db    3258 PVPAVAQDEPLAIVGMACRLPGGVSSPEDLWRLVESGTDAVSGFPTDRGWDVEGLFDPDP 3317

Qy      63 DAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSAL 122
```

```
Db   3318 DAAGKSYRAEGGFLDTAAGFDAGFFGISPREALAMDPQQRLLLEVSWEAFERAGIEPGSV 3377

Qy    123 VGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLG--TMPSVGAGRISYALGLRGPCVA 180

Db   3378 RGSDTGVFIGAFPVGYGAGA---------AREGYGATAAPNVLSGRLSYFFGLEGPAIT 3427

Qy    181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240

Db   3428 MDTACSSSLVALHLAAQALRNGECSMALAGGVTVMATPEVFTEFARQRGLASDGRCKAFA 3487

Qy    241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Db   3488 DSADGAGFSEGAGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGFTAPNGPAQQRVI 3547

Qy    301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360

Db   3548 RQALANAGLTTAEVDVVEAHGTGTTLGDPIEAQAVIATYGQDRE--QPLLLGTLKSNVGH 3605

Qy    361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Db   3606 TQAAAGVSGVIKMVMALQHSTVPRTLHVNEPSRHVDWSAGAVELVTENQSWPVTGRPRRA 3665

Qy    421 GVSSFGMSGTNAHVVLEEAPAA----TCTPPA-PERPAEL--LVLSARTASALDAQAARL 473

Db   3666 GVSAFGVSGTNAHVVLESAPPAQSVNNAQPVATPVVASELVPLVISAKTLPALTEHEDRL 3725

Qy    474 RDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSI 533

Db   3726 RAYLAASPGADMRAVGSTLALTRSVFEHR---------------AVLLGHDTVTVTGTG 3769

Qy    534 ADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEP 593

Db   3770 TAVSNPRVVFVFPGQGWQWLGMGSALRGSSVVFAERMAECAAALSEFVDWDLFAVL---- 3825

Qy    594 ASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLED 653

Db   3826 --DDPAVVDRVDVVQPASWAVMVSLAAVWQADGVRPDAVIGHSQGEIAAACVAGAVSLRD 3883

Qy    654 AVFLVAARGR-LMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAG 712

Db   3884 AARSVTLRSQAIARGLAGRGAMASVALPAHEIEL-----VDGAWIAAHNGPASTVVAGAP 3938

Qy    713 QPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSG 772

Db   3939 EAVDRVLAVHEARGVRVRRIAVDYASHTPHVELIRDELLDITAGIGSQAPVVPWLSTVDG 3998

Qy    773 KACTDEVSSP---GYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPD 829

Db   3999 ----TWVEGPLDVEYWYRNLREPVGFDSAVGQLRAEGDTVFVEVSASPVLLQAM-----D 4049

Qy    830 ARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVP-LPTYPWQRERYW 888

Db   4050 DDVVTVATLRRDDGDATRMLTALAQAFVEGVTVDWPAILGTATTRVPDLPTYAFQHQRFW 4109

Qy    889 IDTK-----------------------------------ADDAARGDRRAPGAGH--- 908

Db   4110 AEGADRSVAGGHPLLGVAVELPESDGVLLTGRVSLATHAWLADHAVRGSVLLPGTGFVEL 4169

Qy    909 -----DEV---------------------------------------------------- 911

Db   4170 VVRAADEVGCDVVDELVIETPLLLPSSASVHLSVSVGEADESGRRGVTVFSRADGADAWT 4229

Qy    912 ------------------------------------------EEGGAVRG-------- 919

Db   4230 RHVSATIGVSGAALSLPELAAWPPAQAQPVGLGDFYDRLTGAGYEYGPAFQGLQAAWRDG 4289

Qy    920 ------------------------------------------------------------ 919

Db   4290 DTVFAEVALAEEQAEEAARFAVHPALLDAALDAGILNTLDNAEQGVRLPFSWNGVQVRAT 4349

Qy    920 ------------------------------------------------------------ 919

Db   4350 GSATLRVADNPTSDGWSVRVADDSGRPVATVDSLVTRPVTADTLGSAADDLLTVVWTEIP 4409

Qy    920 ------------------------------------------------------------ 919

Db   4410 TPQQTGLSVGRFEDLADGDVPVPEVVVCTALPDSSENPLAPLDPPDPLVQTRTLTTQVLQ 4469

Qy    920 -------GDR----------------------RSARLDHP------------------- 930

Db   4470 AVQAWLAGERFTDSTLVVRTGTGLATAGVSGLMRSAQSEHPGRFVLVECDDNLTLQQLAA 4529

Qy    931 -----------------PPESGRREKVEAA-----GDRPFRLEIDEPGVLDHLVL--RVT 966

Db   4530 TVGLDEPRLRVCDGRFEVPRLARANTPESSPLTIPGDRAWLLEQSHSGTLRDLALVPAET 4589

Qy    967 ERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGV 1026

Db   4590 AERPLQSGEVRVDVRAAGLNFRDVLIALGTYPGE--------AVIGAEAAGVVLEVGPEV 4641

Qy   1027 NGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARL 1086

Db   4642 QDLAPGDRVFGLVGGGFGAVAIADRRMLGVIPDGWSFTTAASVPVVFATAYYGLVDLAGL 4701

Qy   1087 QPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLES-LGVRYVSDSRSDRF 1145

Db   4702 SAGESVLIHAAAGGVGMAATQIARHLGARIYATASTGKQHVLREAGLEDARIGDSRTTGF 4761

Qy   1146 VADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGL-RPFLR 1204

Db   4762 REMVLDTTDSRGVDVVLNSLSGDFVDASLDLLPRGGRFVEMGKTDIRDPHQVTADRPGTS 4821

Qy   1205 NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQ 1264

Db   4822 YQAFDLMD------AGPDRLREIIADLLALFAQGVLLPLPVRAWDIRQAREAFSWMSRAR 4875

Qy   1265 HLGKLVLTLGDPEVQIRIPTHAGAGPS---TGDRDLLDRLASAAPAARAAALEAFLRTQV 1321

Db   4876 HIGKIVLT---------VPQQLDANGTVLVTGGSGVL-----AGIAARHLVAEQGVRHLL 4921

Qy   1322 SQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASL-----KLKLSTTFLSTSPNIAL 1376

Db   4922 LLSRSTPD-----DALINELG-------ELGARVDTAICDVSDRAGLARILAGVSPEHPL 4969

Qy   1377 LAQNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEI 1418

Db   4970 TA-----VIHTAGALDDGVVESLTAQQLETVLRPKADGAWHL 5006
```

Search completed: October 2, 2003, 17:35:08
Job time : 93.4261 secs

```
OM protein - protein search, using sw model

Run on:         October  2, 2003, 17:34:03 ; Search time 81.0897 Seconds
                                              (without alignments)
                                              2772.494 Million cell updates/sec

Title:          US-10-014-717-2
Perfect score:  7210
Sequence:       1 VADRPIERAAEDPIAIVGAS..........GVQNDFVSSGADQDWEIIAL 1421

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters:      587654

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
                 1:  /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                 2:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                 3:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                 4:  /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                 5:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                 6:  /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                 7:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                 8:  /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                 9:  /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
                 10:  /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
                 11:  /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
                 12:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                 13:  /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
                 14:  /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
                 15:  /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
                 16:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                 17:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                 18:  /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7210	100.0	1421	14	US-10-014-717-2	Sequence 2, Appli
2	4137	57.4	2439	14	US-10-014-717-7	Sequence 7, Appli
3	3311.5	45.9	7257	14	US-10-014-717-5	Sequence 5, Appli
4	2311.5	32.1	6146	15	US-10-156-761-10436	Sequence 10436, A
5	2249	31.2	6145	15	US-10-156-761-7962	Sequence 7962, Ap
6	2245.5	31.1	3798	14	US-10-014-717-6	Sequence 6, Appli
7	2199.5	30.5	3352	15	US-10-156-761-7961	Sequence 7961, Ap
8	2197.5	30.5	3564	15	US-10-156-761-7964	Sequence 7964, Ap
9	2159.5	30.0	7746	15	US-10-156-761-7965	Sequence 7965, Ap
10	2152	29.8	1835	15	US-10-156-761-7963	Sequence 7963, Ap
11	2149.5	29.8	1784	15	US-10-205-032-18	Sequence 18, Appl
12	2128	29.5	3816	11	US-09-808-880-3	Sequence 3, Appli
13	2125	29.5	5215	10	US-09-861-289-2	Sequence 2, Appli
14	2125	29.5	5215	10	US-09-860-846-2	Sequence 2, Appli
15	2125	29.5	5215	11	US-09-988-384B-2	Sequence 2, Appli
16	2125	29.5	5215	11	US-09-836-821-2	Sequence 2, Appli
17	2114.5	29.3	3745	15	US-10-205-032-14	Sequence 14, Appl
18	2105	29.2	3939	15	US-10-156-761-10434	Sequence 10434, A
19	2077.5	28.8	3613	15	US-10-156-761-10432	Sequence 10432, A
20	2052.5	28.5	3739	10	US-09-861-289-33	Sequence 33, Appl
21	2052.5	28.5	3739	10	US-09-860-846-33	Sequence 33, Appl
22	2052.5	28.5	3739	11	US-09-988-384B-33	Sequence 33, Appl
23	2052.5	28.5	3739	11	US-09-836-821-33	Sequence 33, Appl
24	2052.5	28.5	11877	10	US-09-861-289-6	Sequence 6, Appli
25	2052.5	28.5	11877	10	US-09-860-846-6	Sequence 6, Appli
26	2052.5	28.5	11877	11	US-09-836-821-6	Sequence 6, Appli
27	2052.5	28.5	12199	11	US-09-988-384B-6	Sequence 6, Appli
28	2051.5	28.5	4551	11	US-09-793-708-1	Sequence 1, Appli
29	2051.5	28.5	4551	12	US-10-201-365-2	Sequence 2, Appli
30	2051.5	28.5	4551	12	US-10-160-539-1	Sequence 1, Appli
31	2051.5	28.5	4613	10	US-09-861-289-31	Sequence 31, Appl
32	2051.5	28.5	4613	10	US-09-860-846-31	Sequence 31, Appl
33	2051.5	28.5	4613	11	US-09-988-384B-31	Sequence 31, Appl
34	2051.5	28.5	4613	11	US-09-836-821-31	Sequence 31, Appl
35	2051.5	28.5	4840	15	US-10-156-761-10435	Sequence 10435, A
36	2048.5	28.4	3739	11	US-09-793-708-2	Sequence 2, Appli
37	2048.5	28.4	3739	12	US-10-201-365-3	Sequence 3, Appli
38	2048.5	28.4	3739	12	US-10-160-539-2	Sequence 2, Appli
39	2046.5	28.4	4881	15	US-10-156-761-8481	Sequence 8481, Ap
40	2025	28.1	1925	15	US-10-205-032-12	Sequence 12, Appl
41	2024.5	28.1	3970	15	US-10-156-761-10429	Sequence 10429, A
42	1966.5	27.3	6396	12	US-09-940-316B-72	Sequence 72, Appl
43	1954.5	27.1	1832	14	US-10-014-717-4	Sequence 4, Appli
44	1918.5	26.6	4150	11	US-09-808-880-2	Sequence 2, Appli
45	1889	26.2	3519	11	US-09-808-880-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-10-014-717-2
; Sequence 2, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/10/014,717
;  CURRENT FILING DATE:  2001-11-13
;  PRIOR APPLICATION NUMBER: US/09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1421
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-10-014-717-2

  Query Match            100.0%;  Score 7210;  DB 14;  Length 1421;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1421;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60

Qy         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120

Qy        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180

Qy        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240

Qy        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Qy        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360

Qy        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Qy        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480

Qy        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660

Qy        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720

Qy        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780

Qy        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840

Qy        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900

Qy        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960

Qy        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020

Qy       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080

Qy       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140

Qy       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200

Qy       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260

Qy       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320

Qy       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380

Qy       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
              |||||||||||||||||||||||||||||||||||||||||
Db       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
```

RESULT 2
US-10-014-717-7
; Sequence 7, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 2439
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-10-014-717-7

```
  Query Match             57.4%;  Score 4137;  DB 14;  Length 2439;
  Best Local Similarity   43.4%;  Pred. No. 6.3e-309;
  Matches  933;  Conservative  135;  Mismatches  308;  Indels  776;  Gaps   13;

Qy          6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65
              :|:   :||||||  || |||        || ||:  || |   | :|  |     |  : |
Db         26 LEQERTEPIAIVGIGCRFPGGADTPEAFWELLDSGRDAV--QPLDRRWALVGVHPSEEVP 83

Qy         66 GKTPVTRASFLSD-VACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124
                   |  |:: |  |||:||| |||||  :||  |||||| || ||:| ||| :| |
Db         84 -----RWAGLLTEAVDGFDAAFFGTSPREARSLDPQQRLLLEVTWEGLEDAGIAPQSLDG 138

Qy        125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184
              : |:||:  |   |:|   :      | ||:   |   || |||:||.|||:|||: ||||
Db        139 SRTGVFLGACSSDYSHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPCLTVDTA 197

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244
              |||||:||||:|||:  |    ||||||:::||   |:: |  :  :||: || |:  | | |:
Db        198 CSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHCRTFDASAN 257

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304
              || ||||| :|||||||  |:  ||||  |:||||||:|  || |:||  ||    :||  :|: ||
Db        258 GFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLREAL 317

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364
                |    | ::||||  |||||:||||||::||  ||  |   |    :   ::|:||||||| | |
Db        318 QSARVDAGAIGYVETHGTGTSLGDPIEVEALRAVLGPARADGSRCVLGAVKTNLGHLEGA 377

Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424
              :|: ||:  |:|  |   ||  :||   |||||       |  :          |||      || ||||:
Db        378 AGVAGLIKAALALHHELIPRNLHFHTLNPRIRIEGTALALATEPVPWPRAGRPRFAGVSA 437

Qy        425 FGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQC 484
```

```
Db      438 FGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSAAALDAQAARLSAHIAAYPEQG 497

Qy      485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544

Db      498 LGDVAFSLVSTRSPMEHRLAVAATSREALRSALEVAAQGQTPAGAARGRAASSPGKLAFL 557

Qy      545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604

Db      558 FAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCEVMWAEPGSSRSSLLDQT 617

Qy      605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664

Db      618 AFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAGVFSLEDAVRLVVARGRL 677

Qy      665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724

Db      678 MQALPAGGAMVSIAAPEADVAAAVAPHAALVSIAAVNGPEQVVIAGAEKFVQQIAAAFAA 737

Qy      725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784

Db      738 RGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSIALVSNLSGKPCTDEVSAPGY 797

Qy      785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844

Db      798 WVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACLPDARPVLLPASRAGRDE 857

Qy      845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI----DTKADDAAR-- 898

Db      858 AASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERYWIEAPVDREADGTGRAR 917

Qy      899 ------------------------------------------------------------ 898

Db      918 AGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEVVFPGAGYLEMALSSGAE 977

Qy      899 -------------------------------------------------GDRRAP----- 904

Db      978 ILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQVASREPGERRAPFRIHA 1037

Qy      905 ------------------------------------------------------------ 904

Db     1038 RGVLRRIGRVETPARSNLAALRARLHAAVPAAAIYGALAEMGLQYGPALRGLAELWRGEG 1097

Qy      905 ------------------------------------------------------------ 904

Db     1098 EALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSP 1157

Qy      905 ------------------------------------------------------------ 904

Db     1158 GELWCHARVVSDGQQASSRWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLE 1217

Qy      905 ------------------------------------GAGH-------------------- 908

Db     1218 LDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVVHAAGDDTSTAGMRALLA 1277

Qy      909 ------------------------------------------------------------ 908

Db     1278 NAFDGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDADALESALMRGCDSVLSL 1337

Qy      909 ------------------------------------------------------------ 908

Db     1338 VQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDP 1397

Qy      909 ----------------DEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEI 952

Db     1398 AEPEGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPDAQRREKVEPAGDRPFRLEI 1457

Qy      953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012

Db     1458 DEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALGVAPNDLPGEEIEPLVLG 1517

Qy     1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072

Db     1518 SECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLA 1577

Qy     1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132

Db     1578 YLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAEVYATADTPENRAYLESL 1637

Qy     1133 GVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY 1192

Db     1638 GVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLMVLRACGRLVKLGRRDDC 1697

Qy     1193 ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGV------------ 1239

Db     1698 ADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLVAAGAISPLGSGLRVGGS 1757

Qy     1240 FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHA------------ 1286

Db     1758 LTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRAPAESSVAVRADGTYLVT 1817

Qy     1287 -----------------GAG---------------------------------------- 1289

Db     1818 GGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALEAHGARVTVAKADVADRS 1877

Qy     1290 ------------------------------------------------------------ 1289

Db     1878 QIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARFRTVMGPKVQGALHLHTLTREA 1937

Qy     1290 ------------------------------------------------------------ 1289

Db     1938 PLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAQGLPALSIDWGMFTEVGMAVA 1997

Qy     1290 -----------------------------------------------------PST------- 1292

Db     1998 QENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITPRQWVEFYPATAASRRLS 2057

Qy     1293 ---------------GDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALF 1338

Db     2058 RLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQSHVLRLPEDKIEVDAPL 2117

Qy     1339 TRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL-DALATAL 1389

Db     2118 SSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWLLDDALVVRL 2169
```

RESULT 3
US-10-014-717-5
; Sequence 5, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-10-014-717-5

```
  Query Match             45.9%;  Score 3311.5;  DB 14;  Length 7257;
  Best Local Similarity   41.1%;  Pred. No. 1.2e-244;
  Matches  784;  Conservative  162;  Mismatches  432;  Indels  531;  Gaps   25;

Qy        9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKT 68
```

```
Db   3021 ASDEPIAIVGAACRFPGGVEDLESYWQLLAEGVVVSAEVPADRWDAADWYDPDPEIPGRT 3080

Qy     69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETG 128
Db   3081 YVTKGAFLRDLQRLDATFFRISPREAMSLDPQQRLLLEVSWEALESAGIAPDTLRDSPTG 3140

Qy    129 VFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSS 188
Db   3141 VFVGAGPNEYYTQRLRGFTDGAAGLYGGTGNMLSVTAGRLSFFLGLHGPTLAMDTACSSS 3200

Qy    189 LVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR 248
Db   3201 LVALHLACQSLRLGECDQALVGGVNVLLAPETFVLLSRMRALSPDGRCKTFSADADGYAR 3260

Qy    249 GEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAG 308
Db   3261 GEGCAVVVLKRLRDAQRAGDSILALIRGSAVNHDGPSSGLTVPNGPAQQALLRQALSQAG 3320

Qy    309 CAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
Db   3321 VSPVDVDFVECHGTGTALGDPIEVQALSEVYGPGRSGDRPLVLGAAKANVAHLEAASGLA 3380

Qy    369 GLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMS 428
Db   3381 SLLKAVLALRHEQIPAQPELGELNPHLPWNTLPVAVPRKAVPWGRGARPRRAGVSAFGLS 3440

Qy    429 GTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDV 488
Db   3441 GTNVHVVLEEAPEVEPAPAAPARPVELVVLSAKSAAALDAAAARLSAHLSAHPELSLGDV 3500

Qy    489 AFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQ 548
Db   3501 AFSLATTRSPMEHRLAIATTSREALRGALDAAAQQKTPQGAVRGKAVSSRGKLAFLFTGQ 3560

Qy    549 GAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQ 608
Db   3561 GAQMPGMGRGLYETWPAFREAFDRCVALFDREIDQPLREVMWAAPGLAQAARLDQTAYAQ 3620

Qy    609 PALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQAL 668
Db   3621 PALFALEYALAALWRSWGVEPHVLLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMQAL 3680

Qy    669 PAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGAR 728
Db   3681 PAGGAMVAIAASEAEVAASVAPHAATVSIAAVNGPDAVVIAGAEVQVLALGATFAARGIR 3740

Qy    729 TKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRH 788
Db   3741 TKRLAVSHAFHSPLMDPMLEDFQRVAATIAYRAPDRPVVSNVTGHVAGPEIATPEYWVRH 3800

Qy    789 AREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATV 848
Db   3801 VRSAVRFGDGAKALHAAGAATFVEVGPKPVLLGLLPACLGEADAVLVPSLRADRSECEVV 3860

Qy    849 LEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDA------------ 896
Db   3861 LAALGAWYAWGGALDWKGVFPDGARRVALPMYPWQRERHWMDLTPRSAAPAGIAGRWPLA 3920

Qy    897 ------------------------------------------------------------ 896
Db   3921 GVGLCMPGAVLHHVLSIGPRHQPFLGDHLVFGKVVVPGAFHVAVILSIAAERWPERAIEL 3980

Qy    897 --------------------------ARGD---------------RR------------- 902
Db   3981 TGVEFLKAIAMEPDQEVELHAVLTPEAAGDGYLFELATLAAPETERRWTTHARGRVQPTD 4040

Qy    903 -APGA------------------------------------------------------- 906
Db   4041 GAPGALPRLEVLEDRAIQPLDFAGFLDRLSAVRIGWGPLWRWLQDGRVGDEASLATLVPT 4100

Qy    907 ---GHD---------------------EVEEGGA-------------------VR-GGD 921
Db   4101 YPNAHDVAPLHPILLDNGFAVSLLSTRSEPEDDGTPPLPFAVERVRWWRAPVGRVRCGGV 4160

Qy    922 RRSA------------------------------------------------RLDHPPP--- 932
Db   4161 PRSQAFGVSSFVLVDETGEVVAEVEGFVCRRAPREVFLRQESGASTAALYRLDWPEAPLP 4220

Qy    933 ----------------------------------------------------------ESGR 936
Db   4221 DAPAERIEESWVVVAAPGSEMAAALATRLNRCVLAEPKGLEAALAGVSPAGVICLWEAGA 4280

Qy    937 REKVEAAGDR-------------------------------------------------- 946
Db   4281 HEEAPAAAQRVATEGLSVVQALRDRAVRLWWVTMGAVAVEAGERVQVATAPVWGLGRTVM 4340

Qy    947 ------------------------------------------------------------ 946
Db   4341 QERPELSCTLVDLEPEADAARSADVLLRELGRADDETQVAFRSGKRRVARLVKATTPEGL 4400

Qy    947 ------PFRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDD 1000
Db   4401 LVPDAESYRLEAGQKGTLDQLRLAPAQRRAPGPGEVEIKVTASGLNFRTVLAVLGM---- 4456

Qy   1001 LPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQA 1060
Db   4457 YPGDAGP---MGGDCAGVATAVGQGVRHVAVGDAVMTL--GTLHRFVTVDARLVVRQPAG 4511

Qy   1061 LSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATA 1120
Db   4512 LTPAQAATVPVAFLTAWLALHDLGNLRRGERVLIHAAAGGVGMAAVQIARWIGAEVFATA 4571

Qy   1121 GTPEKRAYLESLGV--RYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLR 1178
Db   4572 -SPSKWAAVQAMGVPRTHIASSRTLEFAETFRQVTGGRGVDVVLNALAGEFVDASLSLLS 4630

Qy   1179 SHGRFVELGKRDCYADNQ--LGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIA 1236
Db   4631 TGGRFLEMGKTD-IRDRAAVAAAHPGVRYRVFDILEL------APDRTREILERVVEGFA 4683

Qy   1237 AGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPT-----HAGAGP- 1290
Db   4684 AGHLRALPVHAFAITKAEAAFRFMAQARHQGKVVL-LPAPSAAPLAPTGTVLLTGGLGAL 4742

Qy   1291 -----------------STGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVG 1333
Db   4743 GLHVARWLAQQGVPHMVLTGRRG-LDTPGAAKAVAEIEALGARVTIAASDVADRNALEAV 4801

Qy   1334 AEAL---------------------------FTRLGMDSLMAVELRNRIEASLKLKLST 1365
Db   4802 LQAIPAEWPLQGVIHAAGALDDGVLDEQTTDRFSRVLAPKVTGAWNLHELTAGNDLAFFV 4861

Qy   1366 TFLSTS--------PNIALLAQNLLDALATALSLERVAAENLRAGVQND 1406
Db   4862 LFSSMSGLLGSAGQSNYA-AANTFLDALAAHRRAEGLAAQSLAWGPWSD 4909
```

RESULT 4
US-10-156-761-10436
; Sequence 10436, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30

```
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10436
;   LENGTH: 6146
;   TYPE: PRT
;  .ORGANISM: Streptomyces avermitilis
US-10-156-761-10436

  Query Match             32.1%;  Score 2311.5;  DB 15;  Length 6146;
  Best Local Similarity   32.3%;  Pred. No. 9.5e-168;
  Matches  635;  Conservative  178;  Mismatches  519;  Indels  631;  Gaps   37;

Qy        5 PIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPD 63
Db     3983 PTAAEADDPVVIVAMACRFPGEATSPEKLWDLIAAGKDGIGDFPADRGWSVAA------D 4036

Qy       64 APGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALV 123
Db     4037 AAFS---QTGGFLPDVAEFDAGFFGISPREALAMDPQQRLLLETSWEALERAGVDALKLR 4093

Qy      124 GTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDT 183
Db     4094 GSRTGVFVGAGSHDYGTLLTSLEGGQDYALTGAVG---SVLSGRIAYVLGLEGPALTVDT 4150

Qy      184 AYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEA 243
Db     4151 ACSSSLVALHLAAQALRGGECDVALAGGVAVMATPDAFDAFARQGGLAPDGRCKAFADGA 4210

Qy      244 DGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRA 303
Db     4211 DGTGWGEGVGVLVLTRLSEARRRGHDVLAVVRGSAVNSDGASNGLTAPNGPSQQRVIRQA 4270

Qy      304 LADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEY 363
Db     4271 LASAGLSAADVDMVEAHGTGTSLGDPIEAQALLATYGQDRPADRPLWLGSVKSNIGHTQA 4330

Qy      364 ASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVS 423
Db     4331 AAGVAGVIKSVLALRNGLLPATLHVDEPSREVDWSAGAVELLTEGREWPETDGPRRAGVS 4390

Qy      424 SFGMSGTNAHVVLEEAPAA--TCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYP 481
Db     4391 AFGISGTNAHVILEQAPSVEEPVTAPADDGLLVPWVVSARSEEALRAQARRLADHVAA-T 4449

Qy      482 SQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKL 541
Db     4450 DLTVADVGYSLAGTRAGLEHRAVLTGREREDFLVQLAALAEGASAAGMVRGVA--GEGGT 4507

Qy      542 AFLFTGQGAQTLGMGRGLYDVWSAFREAFD-LCVRLFNQELDRPLREVMWAEPASVDAAL 600
Db     4508 AFLFTGQGAQRLGMGRELYERFPVFAAAFDGVCAQL-DLLLERPVKEVVF-----TDAPA 4561

Qy      601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
Db     4562 LDRTVFTQAGLFALEVALFELVSSWGVRADVLLGHSIGELAAAYVAGVWSLADACRIVAA 4621

Qy      661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
Db     4622 RGRLMQALPEGGAMVAVEAAEDELPDL----PEGVSVAAVNGPRSLVLSGDEEPVTALAR 4677

Qy      721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
Db     4678 TFAEQGRRTKQLAVSHAFHSARMEPMLAEFAETLAAVEFRSPRIPVVSNVTGGMAGAEFT 4737

Qy      781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTL----------LGLVPACMPDA 830
Db     4738 TPAYWVRHVREAVRFADGVATVLARGVDRFLELGPRGALTAMAEETLDHTGADAVCVPVL 4797

Qy      831 RPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWID 890
Db     4798 HPE--------RPEDATLLHALATASVTGAAVDWTTPF-ACARTTELPTYAFQRRRFW-P 4847

Qy      891 TKADDAARGDRRAPG---AGH--------------------------------------- 908
Db     4848 RPAASAGAGDLAAAGLAEAGHPLLTAWLPSPEGDGVLCTGRLSLATHPWLGDHTVLGTVL 4907

Qy      909 ------------------------------------------------------------ 908
Db     4908 VPGTAFVDLVGWVGDRVGCGGIRELTLATPLALAEDAAVRLRVVLGAPDDTGCRPVAVYA 4967

Qy      909 ------------------------------------------------------------ 908
Db     4968 QPDTEDEGTDWTRHAEGLLAPADASDTPAAPPGPALDAWPVPGAEPLPMDGFYEGLADAG 5027

Qy      909 ----------------DEV--------EEGGA---------------------------- 916
Db     5028 FAYGPLFRGLRAAWRRGDEVFAEVSLPADESGAGFGVHPALLDAALHALGPAAQASDEPG 5087

Qy      917 ---------------------------VRGGD---------------------------- 921
Db     5088 SARLPFSWGEVRVHAVGADLLRVRLVRAADGTVTLDAADAAGRPVVSIGSLVLRPLSPDR 5147

Qy      922 ------------------------------------------------------------ 921
Db     5148 LRAGTQAPDDALFATRWIPLDVTGDTGDTARDTVECVLLGEPSAGAWRSHPDLAAYAAAL 5207

Qy      922 --------------------------RRSAR----------------------------- 926
Db     5208 ADGKETPGPVLARCPLTTGDPAEAARRSAEWTLDLLQQWLADDRLTDSHLVIGTRHAAPV 5267

Qy      927 ------------LDHPPPESGRREKVEAA------------------------------- 943
Db     5268 ATPYAAPVDATDADAAPVDAGHEAPVDTAAAALAQSTLAGLVRSAQTENPARITLVDFDT 5327

Qy      944 --------------------------------------------GDRPFRLEIDEPGVLD 959
Db     5328 TAPDPAHLTAAVRLGEPEVAVRQGTLYARRLTRPETGRALTVPPGPGPWRLDSTGRGSLD 5387

Qy      960 HLVLRVTERRAPGL--GEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAG 1017
Db     5388 HLALVPCPDAAAPLAEGTVRIAVHAAGVNFRDVLIALDM----YPGRAD----LGTECAG 5439

Qy     1018 RIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAW 1077
Db     5440 VVLETGPGVTGLAPGDRVMGMVAGAFGPTAVADRRVLARIPDGWSFETAAATPVAFLTAY 5499

Qy     1078 YALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLES-LGVRY 1136
Db     5500 YGLVDLAGLGAGESVLVHAAAGGVGMAAVQLARHLGAEVFGTASEPKWSTLLDTGLDRAH 5559

Qy     1137 VSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQ 1196
Db     5560 IGSSRTLDFADLVRETTGGAGVDVVLNSLAGEFTDASLRAMPRGGRFVEMGKTDPRDPER 5619

Qy     1197 LGL-RPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVAD 1255
Db     5620 VAADHPGVRYRSFDLGE------AGPDRIAEILAHLVELFESGALTPLPVTTWDIRDATD 5673

Qy     1256 AFRSMAQAQHLGKLVLTL------------------------------------------ 1273
Db     5674 AFRVLSQATLTGKAVLTVPAGSVPFEAGETVLITGGTGTLGTLLARHLVTEHGLRHVTLA 5733

Qy     1274 -----GDPEV------------QIRIPTHAGAGPSTGDRDLLD------RLASAAPAARA 1310
Db     5734 GRRGTDTPEVRQLRADLAEMGAEVRVEA-CDAGDEDAVRRLLEPLTAHHRLAGVVHAAGV 5792

Qy     1311 ---AALEAFLRTQVSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTF 1367
Db     5793 TDDGVVSALDRRRLSAVLH-PKVR-GAWNL---------------HRLTAHLAPRMFVLF 5835

Qy     1368 LSTSPNIALLAQN-------LLDALATALSLERVAAENLRAGV 1403
Db     5836 SSASATLGAAGQGNYAAANAFLDALAEHRHARGLPASSLAWGL 5878
```

```
RESULT 5
US-10-156-761-7962
; Sequence 7962, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7962
;   LENGTH: 6145
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-7962

  Query Match             31.2%;  Score 2249;  DB 15;  Length 6145;
  Best Local Similarity   37.2%;  Pred. No. 6.2e-163;
  Matches  611;  Conservative  183;  Mismatches  545;  Indels  304;  Gaps   50;

Qy          1 VADRPIERAAED-PIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWF 58
Db       1531 VVEQEQHAAADDEPIAIVAMSCHLPGGVDSPEALWDLVTSGGDAISEFPSNRGWDVEALY 1590

Qy         59 DPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIA 118
Db       1591 DPDPDRPGKTYARDGGFLYGATDFDAGFFGISPREALAMDPQQRLMLETSWEAFERAGIA 1650

Qy        119 PSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPC 178
Db       1651 PAQLRGSRTGVFVGMAYQGYGADVRRTPEG--VEGHRLVGGASSVVSGRVAYTFGLEGPA 1708

Qy        179 VAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKA 238
Db       1709 VTIDTACSSSLVALHLAMQSLRNGECTMALAGGVTVMAGPSVFVEFSRQRGLSPDGRCRA 1768

Qy        239 FSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEI 298
Db       1769 FGADADGTGWAEGVGMLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLTAPNGPAQQR 1828

Qy        299 VLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNL 358
Db       1829 VIHQSLASAGLAPRDVDAVEAHGTGTRLGDPIEAQALLATYGQGRDGDRPLWLGSLKSNI 1888

Qy        359 GHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISW--GDLRLTVTRARTPWPDWNT 416
Db       1889 GHTQAAAGVAGVIKMVMAMRHGILPRTLHADEPTPHVDWSAGDVRL-LTEA-VDWPESDR 1946

Qy        417 PRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAEL-LVLSARTASALDAQAARLRD 475
Db       1947 PRRAAISSFGVSGTNAHTIIEQAPATAELPPTPDSGALVPWVLSGKGEPALRAQAARLAD 2006

Qy        476 HLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIAD 535
Db       2007 HLDANDGWTASDIGLSLA-SRETFEDRAVLLAQGSDKLRQALTALAAGQPAANVTLGRA- 2064

Qy        536 SSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFD-LCVRLFNQELDRPLREVMWAEPA 594
Db       2065 RSEGKVGFLFSGQGSQRIGMGRELYEAFPVFADAYDEVCARL--------------EAPV 2110

Qy        595 SVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDA 654
Db       2111 DVDAETLHRTGCAQPALFAVEVALFRLLESWGVRPDYVAGHSVGEIAAAHVAGVLSLDDA 2170

Qy        655 VFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAA-SVSIAAVNAPDQVVIAGAGQ 713
Db       2171 VRLVSARAALMQALPVGGAMVAVQATEEE----VLPHLTDEVGIAAINGPRSMVLSGAGE 2226

Qy        714 PVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGK 773
Db       2227 AVTAIAEVFVGQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFHQPQVPVVSNLTGR 2286

Qy        774 ACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDA--R 831
Db       2287 LA--EPYSPEYWVRHVREAVRFADGIETLGELGVTTFVEIGPGGVLSGMVQDCVDEAVTV 2344

Qy        832 PALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDT 891
Db       2345 PVL----RGDRPEQQAVVTALAELHVHGVSPDWQALFP-GAHRVTLPTYAFQRERFWLDG 2399

Qy        892 KADDAARG--DRRAPGAGHDEVE----EGGAVRGGDRRSARLDH--PPPESGRREKVE-- 941
Db       2400 EGDLAAAGTPTDAADSGFWDSVEREDAESLAATLGVSADASLGAILPRLSAWRRQRREQS 2459

Qy        942 ----------------------------AAGDRPF----RLEIDEPGVLDHLVLRVTER 968
Db       2460 VVDGWRYRVTWKPLGALPQPGASGTWLLVVAGESEWTASVRTALDERGL--ELVTLVA-- 2515

Qy        969 RAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNG 1028
Db       2516 -GPGTDRGTLARELAGV--GPVTGVLSLLAEDESASAGDPGLSHGLAA--TLCLVQALGD 2570

Qy       1029 LVVGQPVIALSAGAFATHVT---------------TSAALVLPR--------PQALSAIE 1065
Db       2571 VGVDVPLWCATRGAVATGRSDRVDRPLQSQVWGFGRAAALEHPRRWGGLLDLPELLDART 2630

Qy       1066 AAAMPVAYLTAWYALDRIA---------RL-------------QPGERVLIHAATGGVGL 1103
Db       2631 AAR--VAAVLGQTAEDQVAVRASGVLGRRLVRAARATGPAQGWSPRGTVLITGGTGALGG 2688

Qy       1104 AAVQWAQHVGAEVH---------ATAGTPEKRAYLESLGVR------------------- 1135
Db       2689 HVARWLAGAGAE-HLVLTSRRGPAAPGAAELVAELEELGAAATVVACDAADRDALRALLA 2747

Qy       1136 -----------------YVSDSRSDRFVADVRA------------------------- 1151
Db       2748 QHPVNAVVHAAGVGDHVMIEDSDPAGFAGSVAAKAAGATHLDELLAGQELDAFVMFSSGA 2807

Qy       1152 --WTG-GEGVDVVLNSLSGELIDKSFNLLRSHGR---------FVELGKRDCYADNQLGL 1199
Db       2808 GIWGGAGQGAYSAANAYLDALAEHR----RAHGRTALAVSWGGWAEGGMAG-VGDGEEML 2862

Qy       1200 RPFLRNLSFSLVDLRGMMLERPARVRALLEELL--GLIAAGV-------FTPPPIATLPI 1250
Db       2863 R-----------RRGLPPMRPALAISALQQALADGETALTVADFDWERFIGPFTVTRPS 2910

Qy       1251 ARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARA 1310
Db       2911 ALFSDV-------------------PEAGRALEAPAPGGAEAGNA-LAAKLTGLPAAEQD 2950

Qy       1311 AALEAFLRTQVSQVL---RTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTF 1367
Db       2951 RTLVDLVRTHAAAVLGHDGTAAVEPG--RAFNDLGFDSLSAVELRNKLTADTGLKLPTTL 3008

Qy       1368 LSTSPNIALLAQ----NLLDALATALSLERVAAE------------NLRAGVQN-----D 1406
Db       3009 VFDYPNATALARFLRAELLGSRGAARQEQRTAAVGDEPIAIVAMSCRLPGGVGSPEDLWQ 3068

Qy       1407 FVSSGA--------DQDWEIIAL 1421
Db       3069 LVTSGGDVISGFPEDRGWNVEAL 3091

RESULT 6
US-10-014-717-6
```

```
; Sequence 6, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/10/014,717
;  CURRENT FILING DATE:  2001-11-13
;  PRIOR APPLICATION NUMBER: US/09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 3798
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-10-014-717-6

  Query Match            31.1%;  Score 2245.5;  DB 14;  Length 3798;
  Best Local Similarity   36.2%;  Pred. No. 5.6e-163;
  Matches  605;  Conservative  191;  Mismatches  495;  Indels  379;  Gaps   43;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKT 68
Db       1519 AADDDIAIVGAACRFPGGDEGLETYWRHLAEGMVVSTEVPADRWRAADWYDPDPEVPGRT 1578

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETG 128
Db       1579 YVAKGAFLRDVRSLDAAFFAISPREAMSLDPQQRLLLEVSWEAIERAGQDPMALRESATG 1638

Qy        129 VFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSS 188
Db       1639 VFVGMIGSEHAERVQGLDDDAAL-LYGTTGNLLSVAAGRLSFFLGLHGPTMTVDTACSSS 1697

Qy        189 LVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR 248
Db       1698 LVALHLACQSLRLGECDQALAGGSSVLLSPRSFVAASRMRLLSPDGRCKTFSAAADGFAR 1757

Qy        249 GEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAG 308
Db       1758 AEGCAVVVLKRLRDAQRDRDPILAVVRSTAINHDGPSSGLTVPSGPAQQALLRQALAQAG 1817

Qy        309 CAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
Db       1818 VAPAEVDFVECHGTGTALGDPIEVQALGAVYGRGRPAERPLWLGAVKANLGHLEAAAGLA 1877

Qy        369 GLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMS 428
Db       1878 GVLKVLLALEHEQIPAQPELDELNPHIPWAELPVAVVRRAVPWPRGARPRRAGVSAFGLS 1937

Qy        429 GTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDV 488
Db       1938 GTNAHVVLEEAPAVEPVAAAPERAAELFVLSAKSAAALDAQAARLRDHLEKHVELGLGDV 1997

Qy        489 AFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIAD-SSRGKLAFLFTG 547
Db       1998 AFSLATTRSAMEHRLAVAASSREALRGALSAAAQGHTPPGAVRGRASGGSAPKVVFVFPG 2057

Qy        548 QGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFT 607
Db       2058 QGSQWVGMGRKLMAEEPVFRAALEGCDRAIEAEAGWSLLGELSADEA---ASQLGRIDVV 2114

Qy        608 QPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQA 667
Db       2115 QPVLFAMEVALSALWRSWGVEPEAVVGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRR 2174

Qy        668 LPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGA 727
Db       2175 ISGQGEMALVELSLEEAEAALRGHEGRLSVAVSNSPRSTVLAGEPAALSEVLAALTAKGV 2234

Qy        728 RTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVR 787
Db       2235 FWRQVKVDVASHSPQVDPLREELIAALGAIRPRAAAVPMRSTVTGGVIAGPELGASYWAD 2294

Qy        788 HAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARP--ALLASSRAGRDEP 845
Db       2295 NLRQPVRFAAAAQALLEGGPALFIEMSPHPILVPPLDEIQTAAEQGGAAVGSLRRGQDER 2354

Qy        846 ATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPG 905
Db       2355 ATLLEALGTLWASGYPVSWARLFPAGGRRVPLPTYPWQHERYWI----EDSVHGSK---- 2406

Qy        906 AGHDEVEEGGAVRGGDRRSARLDHP----------------------------------- 930
Db       2407 ---------PSLRLRQLRNGATDHPLLGAPLLVSARPGAHLWEQALSDERLSYLSEHRVH 2457

Qy        931 ----PPESGRREKVEAAGDRPFRLEIDEPG----VLDHLVL-----------RVTE---- 967
Db       2458 GEAVLPSAAYVEMALAAG-------VDLYGTATLVLEQLALERALAVPSEGGRIVQVALS 2510

Qy        968 RRAPGLGEVEIAV-DAAGLSFNDVQLALGMV---------------------PDDLPGK 1004
Db       2511 EEGPGRASFQVSSREEAGRSW--VRHATGHVCSGQSSAVGALKEAPWEIQRRCPSVLSSE 2568

Qy       1005 PNPPLL-----------------LG-GECAGRIVAVGEGVNGLVVGQPVIALSAGAFAT 1045
Db       2569 ALYPLLNEHALDYGPCFQGVEQVWLGTGEVLGRVRLPGD-----------MASSSGAYRI 2617

Qy       1046 H-------------VTTSAALV------------LPRP-----QALSAI---------- 1064
Db       2618 HPALLDACFQVLTALLTTPESIEIRRRLTDLHEPDLPRSRAPVNQAVSDTWLWDAALDGG 2677

Qy       1065 --EAAAMPV-----AYLTAWYALDRIAR-------------LQPGERVLIHAATGGVGLA 1104
Db       2678 RRQSASVPVDLVLGSFHAKWEVMERLAQAYIIGTLRIWNVFCAAGERHTIDELLVRLQIS 2737

Qy       1105 AV------QWAQH-------VGAEVHATAG----TPEKRAYLESLGVRYVSDSRSDRFVA 1147
Db       2738 VVYRKVIKRWMEHLVAIGILVGDGEHFVSSQPLPEPDLAAVLEEAG----------RVFA 2787

Qy       1148 DVRA---WT--GGEGVDVVL------------------------------NSLSGELI 1170
Db       2788 DLPVLFEWCKFAGERLADVLTGKTLALEILFPGGSFDMAERIYRDSPIARYSNGIVRGVV 2847

Qy       1171 DKSFNLLRSHGRF--VELG----------------KRDCY------------ADNQLGLR 1200
Db       2848 ESAARVVAPSGMFSILEIGAGTGATTAAVLPVLLPDRTEYHFTDVSPLFLARAEQRFRDY 2907

Qy       1201 PFLRNLSFSLVD------------------LRGMMLERPARVRALLEELLGLIAAGVF-- 1240
Db       2908 PFLK---YGILDVDQEPAGQGYAHQRFDVIVAANVIHATRDIRATAKRLLSLLAPGGLLV 2964

Qy       1241 ----TPPPI-----------------------ATLPIARVADAFRSMAQAQHL-----GK 1268
Db       2965 LVEGTGHPIWFDITTGLIEGWQKYEDDLRIDHPLLPARTWCDVLRRVGFADAVSLPGDGS 3024

Qy       1269 LVLTLGDPEVQIRIPTHAGAG-PSTGDRDLLDRLASAAPAARAA--------------A 1312
Db       3025 PAGILGQHVILSRAPGIAGAACDSSGES------ATESPAARAVRQEWADGSADVVHRMA 3078

Qy       1313 LEA--FLRTQVSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLK 1360
Db       3079 LERMYFHRRPGRQVWVHGRLRTGGGA-FTKALAGDLLLFEDTGQVVAEVQ 3127

RESULT 7
US-10-156-761-7961
; Sequence 7961, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
```

```
; APPLICANT:  IKEDA, HARUO
; APPLICANT:  ISHIKAWA, JUN
; APPLICANT:  HORIKAWA, HIROSHI
; APPLICANT:  SHIBA, TADAYOSHI
; APPLICANT:  SAKAKI, YOSHIYUKI
; APPLICANT:  HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE:  2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7961
;   LENGTH: 3352
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-7961

  Query Match            30.5%;  Score 2199.5;  DB 15;  Length 3352;
  Best Local Similarity  37.2%;  Pred. No. 1.6e-159;
  Matches  611;  Conservative  182;  Mismatches  529;  Indels  321;  Gaps   51;

Qy          4 RPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDP 62
Db         27 REVEEKDDDPIAIVGMACRYPGGVRSPEDLWHLVAAGGDAVGDFPADRGWDVEGIYDPDP 86

Qy         63 DAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSAL 122
Db         87 DARGKTYARRGGFLYEAGEFDPGFFGISPREAVAMDPQQRLLLETTWETFERAGIDAESV 146

Qy        123 VGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVD 182
Db        147 RGSRTGVFVGSGYQDYVNLL--VGVGGESDGHLGTGNSASVMSGRIAYTFGLEGPAVTVD 204

Qy        183 TAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAE 242
Db        205 TACSSSLVALHWAIQALRNGECSMALAGGVQVMTTPTAFVEFSRQRGLAPDGRCKAFGAG 264

Qy        243 ADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKR 302
Db        265 ADGTGWAEGVGMLLVERLSDAVRNGHEVLAVVRGSAVNQDGASNGLTAPNGPAQQRVIRQ 324

Qy        303 ALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPE 362
Db        325 ALASAGLSPAQIDVVEAHGTGTTLGDPIEAQALLATYGRERTEDRPLWLGSLKSNIGHSQ 384

Qy        363 YASGITGLLKVVLSLQHGQIPAHLHAQALNPRISW--GDLRLTVTRARTPWPDWNTPRRA 420
Db        385 AAAGVGGIIKMVMAMRHGVLPQTLHVDEPTPHVDWSAGAVRL-LTEA-VEWPETHHPRRA 442

Qy        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELL-------VLSARTASALDAQAARL 473
Db        443 AVSSFGVSGTNAHTIIEQAP-----PAQDEGPAGALVDGPLAWVLSARSDAALRAQAARL 497

Qy        474 RDHL--ETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVR 531
Db        498 RPSARGEVRPQ----DIGLSLATTRAAMRHRAAVVGDNREELLSGLEDLATG--TPSSRV 551

Qy        532 SIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFD-LCVRLFNQELDRPLREVMW 590
Db        552 LLGRPAGGKTGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCAHL-------------- 597

Qy        591 AEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFS 650
Db        598 DAPVDVDAETLHRTGGTQPALFAVEVALFRLLESLGVRPDFVVGHSVGEIAAAHVAGALS 657

Qy        651 LEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAA-SVSIAAVNAPDQVVIA 709
Db        658 LDDAAKLVSARAALMQALPAGGAMVAVQATEEE----VLPHLTDGVSIAAVNGPSSVVVS 713

Qy        710 GAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSN 769
Db        714 GDETAALAIAAGFAEQGRKTSRLKVSHAFHSPLMDPMLAEFAEVVRGLAFEQPQLPVVSN 773

Qy        770 LSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPD 829
Db        774 LTGQPV--EAYTPEYWVRHVRETVRFADSVRTLHDLGVKTFIEIGPGGVLSGMAQGCADD 831

Qy        830 --ARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERY 887
Db        832 IVTVPVL----RADRPERQAVVTALAHLHTHGVSVDWHSFF-AGARRTDLPTYAFQREWY 886

Qy        888 WIDTKADDAA-------------RGDRRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPE 933
Db        887 WVDAPESAAAEADPVDAEFWETVERGDLQA-------LAETLDIGAGDTFGDVL--PRLS 937

Qy        934 SGRREKVE--AAGDRPFRL------EIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGL 985
Db        938 SWRRQRKEQSTVDDRRYRESWKRLGELGPAGIGGTWLLAVPAEENEQTAAVRTALTARGA 997

Qy        986 SFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVV-GQPVI-------- 1036
Db        998 TLKT------LVVDAASHRAG----LAGELAG--IGPVDGVLSLLVNGDPVLPTLLLVQA 1045

Qy       1037 -----------ALSAGAFATH--------------VTTSAALVLPR--------PQAL- 1061
Db       1046 LGDAGVDAPLWCLTSGAVAVSGSDAVRDARHAQVWGLGRTVALELPRRWGGLIDLPESVD 1105

Qy       1062 --SAIEAAAMPVAYLTAWYALDRIA---------------------RLQPGERVLIHAAT 1098
Db       1106 DQAAVRLADVLEQALTGRWEEDQLAVRASGVFARRLAHAPARAATRRWQPRGTVLITGGT 1165

Qy       1099 GGVGLAAVQWAQHVGAEVHA---------TAGTPEKRAYLESLGVR--YVSDSRSDR--- 1144
Db       1166 GALGGHVARWLARGGAE-HLVLTSRRGADAPGAAALRDELEVLGARVTFAACDVADRDAV 1224

Qy       1145 -----------------------------FVADVRAWTGGEGVDVVLNSLSGELIDKS 1173
Db       1225 AALLAQHVFTAVVHAAGVADAGTVDATTPAAFAAALAAKVGGA---AHLDELLGDQELDA 1281

Qy       1174 FNLLRSHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELL- 1232
Db       1282 FVLFSSISGVWGSGGQAAYAAG----NAFLDGLA------------RQRRDRGLTATAVS 1325

Qy       1233 -------GLIAAG------------VFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTL 1273
Db       1326 WGPWADGGMVADGDDEERLRRRGLRAMTP-----------ASAITSLQRALDRDETLLTV 1374

Qy       1274 GD----------------------PEVQ--IRIPTHAGAGPSTGDRDLLDRLASAAPAAR 1309
Db       1375 ADVDWSRFIVPFTLGRPSPLLGDLPEVRAALAVDTPA-AGHGSG--ALAESLAALSHQDR 1431

Qy       1310 AAALEAFLRTQVSQVL-RTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFL 1368
Db       1432 ARALVDLVRTHAAAVLGHRGAGAVEADRPFRDIGFDSLTAVELRNRINKATGLSLPATLI 1491

Qy       1369 STSPNIALLAQNLLDALATALSLERVAAENLRA-----------------GVQN-----D 1406
Db       1492 FDHPTAADLAAHLLAEL-TGGQHEAVAAAATAAVLDEPIAIIGMACRYPGGVRSPEDLWE 1550

Qy       1407 FVSSGAD--------QDWEIIAL 1421
Db       1551 LVASGRDAISRFPGNRGWDVEAL 1573

RESULT 8
US-10-156-761-7964
; Sequence 7964, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
```

```
; APPLICANT:  HORIKAWA, HIROSHI
; APPLICANT:  SHIBA, TADAYOSHI
; APPLICANT:  SAKAKI, YOSHIYUKI
; APPLICANT:  HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE:  2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7964
;   LENGTH: 3564
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-7964

  Query Match             30.5%;  Score 2197.5;  DB 15;  Length 3564;
  Best Local Similarity   34.7%;  Pred. No. 2.5e-159;
  Matches  638;  Conservative  187;  Mismatches  546;  Indels  467;  Gaps   57;

Qy          4 RPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDP 62
Db         26 RDYEDAAGEPVAIIGMSCRYPGGVDSPEDLWELVAGGRDAVSEFPADRGWDLDALYDPDP 85

Qy         63 DAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSAL 122
Db         86 ESRGTSYAREGAFLRDLARFDAGLFGISPREALAMDPQQRLLLETSWELFERAGIDAASL 145

Qy        123 VGTETGVFIGIGPSEYEAAL---PQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCV 179
Db        146 AGSRTGVFAGVMNHEFLAALQNSPQ-----DLEGYLGTGTSGSVASGRVAYTFGLEGPAV 200

Qy        180 AVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAF 239
Db        201 TVDTACSSSLVALHLAVQSLRNGECSLAVAGGVTAMVGPATFVGFSRQRGLAPDGRCKPF 260

Qy        240 SAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIV 299
Db        261 AAGADGTGWGEGAGLLLVERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPSQQRV 320

Qy        300 LKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLG 359
Db        321 IRQALAAAQLTTAQVDAVEAHGTGTRLGDPIEAQALLATYGQGRPDDKPLWLGSIKSNIG 380

Qy        360 HPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISW--GDLRLTVTRARTPWPDWNTP 417
Db        381 HTQAAAGVAGIIKMVMAMRHGELPRTLHVDAPSVHVDWTAGAVRLLT--ERTPWSASDQP 438

Qy        418 RRAGVSSFGMSGTNAHVVLEEAPAA------TCTPPAPERPAELL--------------- 456
Db        439 RRAGISSFGISGTNAHTIIEEAPSAEAFAQDTGAPEGPEEGAESAEPVASEAARPAAGGP 498

Qy        457 ---VLSARTASALDAQAARL--RDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSRE 511
Db        499 VPWVISGRTEEALRAQAERLLSRTDAEVPPL----DVAFSLATTRTALEHRAVLQGETPA 554

Qy        512 GLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFD 571
Db        555 ELAEGLRAVASGSPARGVARGLSGTG-GRVGFLFSGQGSQRLGMGRELYASYPAFAAAYD 613

Qy        572 -LCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPE 630
Db        614 EVCAQL--------------DAVVDVDSEELHRTGAAQPALFAVEVALFRLLESWGIRAD 659

Qy        631 LVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAP 690
Db        660 YVAGHSVGELAAAHVAGVLSLDDAAKLVSARARLMQALPAGGAMVAVQATEDE----VLP 715

Qy        691 HAA-SVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEA 749
Db        716 HLTDAVGIAAINGPRSVVVSGAEDAVTAIAEAFRQQGRKSSRLKVSHAFHSPLMDPILEE 775

Qy        750 FGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGT 809
Db        776 FAEVVRSLTFHRPQIPVVSNLTGRLA--EPYTPQYWVRHVREAVRFADGVRTLNDLGVTT 833

Qy        810 FVEVGPKSTLLGLVPACMPD---ARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAG 866
Db        834 FVEIGPGGVLSALAEGCLDDDVVTVPAL----RADRPERVALVAAVSELHAHGVSPDWHA 889

Qy        867 LFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAG---HDEVEEGGAVRGGDRR 923
Db        890 FFP-GARRVDLPTYAFQYERYWLD--ASPAAPGGVRAAGLGSADHPLLGAAVSLAGGDER 946

Qy        924 --SARLD---HP-------------------------------------------PPES 934
Db        947 LLTGRLSLRTHPWLSDHAVLGTALLPGTAFVELAVRAADEAGCALLEDLTLEAPLVVPEQ 1006

Qy        935 GRRE------KVEAAGDRPFRL------EIDEPGVLDHLVLRVTERRA----------- 970
Db       1007 GGVAVQVWVGAVDGTGRRPLTVHSRPEDDSDLPWV-RHATGFVTEGSEEAGAAAADRTLT 1065

Qy        971 --PGLGEVEIAVDAAGLSFNDVQLALGM----------------------VPD--DLP 1002
Db       1066 AWPPTGSEPVDLD----GFYDRLAALGLAYGPAFRGLRSAWRTGEDVFAEVALPDGTDTG 1121

Qy       1003 GKPNPPLLL--------GGECAGRIVAVGEG-------------------------- 1025
Db       1122 SFLLHPALLDAALHAIGAGGADGSLVAEADGPLLPFAWSRVSVRATGASTLRVRLAPAGT 1181

Qy       1026 --VNGLV---VGQPVI--------ALSA--------------------------GAFAT 1045
Db       1182 DAVSLLVADGAGEPVASVESLTLRAVSAERLRKRSGDAMFTVERAPLTLPSDDADGTAVA 1241

Qy       1046 HVTTSAAL------------VLPRPQA------LSAIEAAAMPVAYLT-AWYALDRIARL 1086
Db       1242 YVPDLAALAEADGPSQPDVVVVPCPDGPEGATEAERVRAVSTDVLRLVQRWFAEDRTPRL 1301

Qy       1087 QPGERV--LIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYL--------------- 1129
Db       1302 VLMARCDDLAHAAAGGLVRSA--QAENPGRIVLLETDRPDEAAALVPGVVLSGEPHVVVR 1359

Qy       1130 ------------ESLGVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSF--N 1175
Db       1360 EGEVSAPRLARAASLATEDAVTGRADEAASALVEGAASRLGTVLLTGAAGAL-GKTLARH 1418

Qy       1176 LLRSHG---------------------RFVELGKRDCYADNQLGLRPFLRNLSF----- 1208
Db       1419 LVTGHGVRRLLLVSRRGADAPGAADLKAELAALGAEATWAACDLAERDALARLLAATPVD 1478

Qy       1209 SLVDLRGMMLE------RPARVRAL----------LEELLG--------LIAAGVFTPPP 1244
Db       1479 SVVHTAGVLDDGVIAALTPERVGAVLAPKADAVLNLDELTGPSTTFVLFSSAAGVFGNPG 1538

Qy       1245 IATLPIARV-ADAFRSMAQAQ-------------------------------------- 1264
Db       1539 QGNYAAANAFLDAFARHRHAQGRPTVSLAWGLWEQEGAMADADEAGRSRMARSGVLPLP 1598

Qy       1265 -----HLGKLVLTLGDPE-VQIRIPTHA---------------------GAGPSTGDRD 1296
Db       1599 TEDGLRLFDAALASNEPVLVPVRLDTAALRNRAAETVPALLRGLVRVSLRGTAQRAAAGS 1658

Qy       1297 LLDRLASAAPAARAAALEAFLRTQVSQVL-RTPEIKVGAEALFTRLGMDSLMAVELRNRI 1355
Db       1659 LAERLAGLAADERDEALLELVRGEVAAVLGHSSSRAVQPTYAFQDLGFDSLTAVELRNRL 1718

Qy       1356 EASLKLKLSTTFLSTSPNIALLAQNL-----------LDALATALSLERVAAE-------- 1397
Db       1719 ATATGLRLPATLVFDHPTPAALARHIGTELLGEAPAVTARAAARQTATAADEPIAIVGMS 1778

Qy       1398 -NLRAGVQND-----FVSSGAD--------QDWEIIAL 1421
Db       1779 CRLPGGVRSPEDLWRLVASGRDGITRFPDNRGWDVDGL 1816
```

```
RESULT 9
US-10-156-761-7965
; Sequence 7965, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7965
;   LENGTH: 7746
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-7965

  Query Match             30.0%;  Score 2159.5;  DB 15;  Length 7746;
  Best Local Similarity   35.0%;  Pred. No. 6.8e-156;
  Matches  626;  Conservative  181;  Mismatches  532;  Indels  451;  Gaps   56;

Qy         11 EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGKTP 69
Db       5971 DEAIAIVGMACRFPGDVRSPEDLWRLVASGGDAISGFPEDRGWDVENLYDPDPDRSGKSY 6030

Qy         70 VTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGV 129
Db       6031 VRHGGFLHEAAEFDPAFFGISPREALAMDPQQRLLLETSWEAFERAGIDPATVRGSRTGV 6090

Qy        130 FIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSSL 189
Db       6091 FAGVMYHDYGGRV--KTAPEGMDAYLGSGSAGSIASGRVSYTFGLEGPAVTVDTACSSSL 6148

Qy        190 VAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGRG 249
Db       6149 VALHLATQALRSGECSMALVGGVTVMATPSTFVEFSRQRGLSADGRCKAFAAGADGTGWA 6208

Qy        250 EGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAGC 309
Db       6209 EGAGMLLVERLSDARRNGHTVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQALANAGV 6268

Qy        310 AASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITG 369
Db       6269 SSEQVDAVEAHGTGTKLGDPIEAQALIATYGQERPGDRPLWLGSLKSNIGHAQAAAGVGG 6328

Qy        370 LLKVVLSLQHGQIPAHLHAQALNPRISW--GDLRLTVTRARTPWPDWNTPRRAGVSSFGM 427
Db       6329 VIKMVMAMRHGALPRTLHVDEPTPHVDWSAGDVRL-LTEA-MEWPETDHPRRAAVSSFGI 6386

Qy        428 SGTNAHVVLEEAPAATCTPPAPER-PAELLVLSARTASALDAQAARLRDHLETYPSQCL- 485
Db       6387 SGTNAHVIVEQAPDTAELPPVPDSGKAVPWVLSGRTEAALRAQAERL---LSLVPEMDLP 6443

Qy        486 -GDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTS-------PGAVRSIADSS 537
Db       6444 PADIGFSLATTRSAMEHRAAVVGESRDELLEGLRALAAGSPSTRVVLGEPGA-------- 6495

Qy        538 RGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVD 597
Db       6496 GGKSGFLFSGQGSQRIGMGRELYDAFPVFAAAYDEVCALVD------------APVDVD 6542

Qy        598 AALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFL 657
Db       6543 SEELNQTGTTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSLEDAAKL 6602

Qy        658 VAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAA-SVSIAAVNAPDQVVIAGAGQPVH 716
Db       6603 VSARAALMQALPADGAMVAVQATEDE----VLPHLTDGVGIAAVNGPQSVVVSGAEDVVV 6658

Qy        717 AIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACT 776
Db       6659 AIGEAFRERGRKTSRLKVSHAFHSPLMDPMLEEFAEAVSGLSFNEPRLPVVSNLTGRLA- 6717

Qy        777 DEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLA 836
Db       6718 -EPYTPEYWVRHVREAVRFADGLHTLHELGVTTFVEIGPGGVLSALAQGCL-DGDIVTVP 6775

Qy        837 SSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDA 896
Db       6776 AVRTDRPEPQAVVTALAELHAHGVSPDWRALFP-GARRVDLPTYPFQYERLWLE--APEE 6832

Qy        897 ARGDRRAPGAGHDEVEE---GGAV----RGGDRRSARLD---HP-----------PPES 934
Db       6833 YTGTAAATGLGLGAAEHPLAGAAVALPGAGGFLVTGRLSLRTHPWLADHTVMDSVLLPGT 6892

Qy        935 GRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDA------AGLSFN 988
Db       6893 ALVELAVRAGEEAGCAQVEDLTLEAPLI--VPER---GGVAVQVWVEAEEEPGRCALSVH 6947

Qy        989 D----------VQLALGMVPDDLPGKPN-------PPLLLGGECAGRIVAVGEGVNGLV 1030
Db       6948 SRREDAPDDAPWVRHAVGRLTDALPEPPHGADLGAWPP---AGAEAVDLTGFYEGLDGLG 7004

Qy       1031 VGQ-PV----------------IAL----SAGAFATH--------------------- 1046
Db       7005 LGYGPVFRGLRSVWRSGDDVLAEVALPEGTEAGAFAVHPALLDAALHAVGAGGLVPVLDG 7064

Qy       1047 ------------------------------------------VTTSAALVLP------- 1056
Db       7065 PLLPFAWSEVGVRAAGATALRVKLSRTGTDAVSLTVADASGGLVATVGSLSLRPVSREQL 7124

Qy       1057 -----------------RPQALSAIEAA-----------------AMP-VAYLTA-- 1076
Db       7125 LAAGGSGRRADDSLFGLAWQPAVLTETDAAAGLDVRSYADLAALSTEPALPDVAVVTVVP 7184

Qy       1077 -------------------------WYALDRIARL----QPGERVLIHAATGGVGLAAV 1106
Db       7185 CPAGSGRETADRVHSTTSEVLALVQWWLAEERAGRLALVTRPGD--LAHAAVWGL----V 7238

Qy       1107 QWAQH--------VGAE--------------------------------VHAT------- 1119
Db       7239 RSAQSENPDRIVLVEAEETDEAVRVLPAALASGEPQFAVRDGEVFVPRLVKATRSATGTA 7298

Qy       1120 AGTPE----------KRAYLESLGVRYVSDSRSDRFV----------------ADVRAW 1152
Db       7299 TGTPDFGVGPVLLTGASGALGGLVARHLVAEHGVRSLLLLSRRGAEAPGAVELEAELAAW 7358

Qy       1153 -----------TGGEGVDVVLNSLS-------------GELIDKSFNLLRSHGRF---- 1183
Db       7359 GAEVRWAACDVADREAVSGMLNGLGEQSLSAVVHTAGVLGDGIVASLTPERMREVFRPKV 7418

Qy       1184 -VELGKRDCYADNQLGLRPFL------------------------------RNLSFSLV 1211
Db       7419 DAVLNLHECTRD--MGLAAFVVFSSVAGMVGSAGQASYAAANSFLDAFSAHRRREGLPAI 7476

Qy       1212 DLR-GMMLERPARVRALLE-ELLGLIAAGVFTPPPIATL---------------PIARVA 1254
Db       7477 SLAWGVWEQSGAMTDGLVEADRARMARSGVLPLPPEEGLRLFDAALASDEAVLAPVRIDT 7536

Qy       1255 DAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALE 1314
Db       7537 GALRAGEAPPVLRALV-----PAAAARRTARAAAAPSAS--SLAERLEEMPEAEREKTVL 7589

Qy       1315 AFLRTQVSQVL-RTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPN 1373
```

```
Db       7590 DLVRAEVAAGLGHASDRTVRPEHAFQDLGFDSLTAVELRNRLNKATGLRLPATLVFDHPT 7649

Qy       1374 IALLAQNLLDALATA----------LSLERVAAE---NLRAGVQNDFVSS 1410

Db       7650 PALLARHVFMAAAGAAEPALVDSLLADLDRVEQELVTTLAASEARDRILS 7699
```

RESULT 10
US-10-156-761-7963
; Sequence 7963, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7963
; LENGTH: 1835
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7963

```
  Query Match             29.8%;  Score 2152;  DB 15;  Length 1835;
  Best Local Similarity   34.1%;  Pred. No. 3e-156;
  Matches  609;  Conservative  185;  Mismatches  554;  Indels  436;  Gaps   49;

Qy          4 RPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDP 62

Db         25 RELEARAQEPIAIVGMSCRFPGGVGSPEELWRLLAEGEDAISQFPVDRGWDVDALYDPDP 84

Qy         63 DAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSAL 122

Db         85 EQPGKVYTRQGGFLHDAADFDAEFFGISPREALAMDPQQRLLLEAAWEAFEHAGIDPAGL 144

Qy        123 VGTETGVFIGIGPSEY---EAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCV 179

Db        145 KGSRTGVFAGLMYHDYGFTPGVLPEG-----VEGLMSTGNSGSVASGRIAYTFGFEGPAV 199

Qy        180 AVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAF 239

Db        200 TVDTACSSSLVALHLAVQALRTGECSMALAGGVTVMAGPSAFVEFSRQRGLAPDGRCKSF 259

Qy        240 SAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIV 299

Db        260 AAAADGASWSEGTGLLLVERLRDARRLGHKVLAVVRGSAVNQDGASNGLTAPNGPSQQRV 319

Qy        300 LKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLG 359

Db        320 ILEALAGARLSAEEVDAVEAHGTGTSLGDPIEAQALLATYGQERE--RPLLLGSLKSNIG 377

Qy        360 HPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRR 419

Db        378 HTQAAAGVAGVIKMVLAMQHGVLPKTLHVDRPTPRVDWSVGSVDLLTEAIAWPETDHPRR 437

Qy        420 AGVSSFGMSGTNAHVVLEEAPAATCTPPAPER------PAELL---------VLSARTAS 464

Db        438 AGVSSFGISGTNAHIILEQAPEPEAGDPADDTAEGEGVAEELLRARTTVVPWVLSAKNGA 497

Qy        465 ALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQ 524

Db        498 ALREQAQRLLSSVD-LAAASPADVGFSLAATRSALEHRAAVVGADGEELRAILTALAAGE 556

Qy        525 TSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAF-DLCVRLFNQELDR 583

Db        557 PAAQVVTGRAGADRGKVGFLFSGQGSQRIGMGRELYEAYPVFAAAYNEVCALL---DL-- 611

Qy        584 PLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAA 643

Db        612 ---------PVDVDAETLHQTGSTQPALFAMEVALFRLLESWAIRPDYVAGHSVGEIAAA 662

Qy        644 CVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPH-AASVSIAAVNA 702

Db        663 HVAGVLSLGDAARLVSVRAALMQALPAGGAMVAVQAAEDE----VLPYLTGEVGIAAING 718

Qy        703 PDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRP 762

Db        719 PQSVVVSGAEDAVMAVAEVFAAQGRKTSRLKVSHAFHSPLMDPMLEEFAAVVRGLTFGEP 778

Qy        763 SIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGL 822

Db        779 RIPVVSNLTGRLA--EPYTPEHWVRHVREAVRFADGVRTLHELGVTTFVEVGPGGVLSAL 836

Qy        823 VPACMPDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPW 882

Db        837 AQGCLDDADIVTVPALRTDRPEPQALTLAVGQLHTRGVSPDWQALFP-GALRVDLPTYAF 895

Qy        883 QRERYWI--DTKADDAARG----------------------DRRAPG---------- 905

Db        896 QRERYWLKGTTAVDVSSAGLDSSGHALLGAVVETAGAQDVLFTGRLSPADLPWIAEHTVG 955

Qy        906 -------------------AGHDEVE----------EGGAVR---------GGDRRSA 925

Db        956 GVPVLPVAAFVELALSAGRQAGFPYLEQLDVDAPLVLAEQGAVRIQVQIGADDDPDRRSV 1015

Qy        926 RL-DHP-----------------------------PPESGRREKVEA---------AG 944

Db       1016 RVFSRPAAGSSGEPWVRHVTGKLTRNAPAPEFDQMIWPPTGAEPGDVEGRPGLRAAWRAG 1075

Qy        945 DRPP---RLEIDE----------PGVLDHLVLRVTERRA-------------------PG 972

Db       1076 DELYAEAELPVTERDRADRYALHPVLLDAALGMATSDRAADPGEVLMPHAWSGMSVHTPG 1135

Qy        973 LGEVEIAV-----DAAGLSFNDVQ-------------------LALGMVPDD------- 1000

Db       1136 PGAVRIRIRPVGDDAVAIELSDADGSPVASVASVRTRTMTAQQVLASRVAHQDWMFEIDW 1195

Qy       1001 -------------------------------------------------LPGKPNPPLL 1010

Db       1196 VEHTAASAPATAAVADWAVLGDEGVHSDIDALRKAVDGGAATPAHVLLPCAPSEGDAAAA 1255

Qy       1011 LGGECAGRIVAVGEGVNG--------LVVGQPVIALSAGAF-ATHVTTSA---ALVLPRP 1058

Db       1256 TRGALAGVLSVVQEWVADERFSDARLVVMTRGAVELDGTALDLAHVAVSGLVKSAQTEHP 1315

Qy       1059 QALSAIEA------------AAMPVAYLTAWYALD-------------RIARLQPGER- 1091

Db       1316 DRLLLVDADDSVTNGGTVTDGGVDPALAAALAALDEPQAAVRDGKLYVPRLARITPPEEP 1375

Qy       1092 ----------VLIHAATGGVGLAAVQWAQHVGAE------------VHATAGTPEKRAYL 1129

Db       1376 TGHAWDPEGTVLLTGATGGLGRSL---ARHLVAERGARRLLLASRRGPAAEGVGELVAAL 1432

Qy       1130 ESLG--VRYVSDSRSDR-FVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVEL 1186

Db       1433 SELGAHVDVAACDIADRDALAELLASVPAEHPLVAVVHAAAALDDGVITALTP------- 1485

Qy       1187 GKRDCY----ADNQLGLRPFLRNLSFS----LVDLRGMM-----------------LERP 1221

Db       1486 GRLDTVLRPKADGALHLHELTRDLNLSAFVLFSSLAGVLGSAGAAGYAAANAFLDGLARQ 1545

Qy       1222 ARVRAL-------------------------------------LEELLGLIAAGVFTPPP 1244

Db       1546 RRAAGLPGTSLAWGLWTDNGGMGDRITDADRNRMGRAGVVPFSPAEGLALFDAALTVDRP 1605
```

```
Qy       1245 IATLPIARVADAFRSMAQAQHLGKLVLTL-GDPEVQIRIPTHAGAGPSTGDRDLLDRLAS 1303
Db       1606 V-IVPVRLDVPGLRAMARAGMLPAVLRGLVGDTGAA------AGEAAAAGTTALARRLAG 1658

Qy       1304 AAPAARAAALEAFLRTQVSQVL-RTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLK 1362
Db       1659 IPEADQEKVMLDLVRHQAAAALGHATAARIEPTTAFRDIGFDSLTAVELRNRLKAATGLR 1718

Qy       1363 LSTTFLSTSPNIALLAQNL-----LDALATAL----SLERVAAE 1397
Db       1719 LKPTLVFDHPNPVAIARLLHKDLVLDLTAEAPPVLGELDRLEAQ 1762
```

RESULT 11
US-10-205-032-18
; Sequence 18, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-2US
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
;   LENGTH: 1784
;   TYPE: PRT
;   ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-18

```
  Query Match             29.8%;  Score 2149.5;  DB 15;  Length 1784;
  Best Local Similarity   37.4%;  Pred. No. 4.5e-156;
  Matches  577;  Conservative  168;  Mismatches  555;  Indels  241;  Gaps   37;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGK 67
Db         32 ALHGPIAVVGMACRYPGGVSSPEDLWRLVETGTDAIGGFPTDRGWDVDAVYDPDPESRNT 91

Qy         68 TPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTET 127
Db         92 TYCREGGFLAGAGDFDAAFFGVSPHEAVVMDPQQRLLLEVSWEALERSGTDPHSLRGSRT 151

Qy        128 GVFIGIGPSEYEAALPQATASAEIDAHGGL---GTMPSVGAGRISYALGLRGPCVAVDTA 184
Db        152 GVYVGAAHQGYAVDAGQVPEGAE-----GFRLTGSADAVLSGRISYLLGLEGPALTVETA 206

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244
Db        207 CSSSLVAVHLAVQALRRGECGLALAGGVAVMPDPAAFVEFSRQRGLAADGRCRAFGAGAD 266

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304
Db        267 GTGWAEGVGVLVLQRLSDAVRDGRWVLGVIRGSAVNQDGASNGLTAPSGPAQQRVIRQAL 326

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364
Db        327 TDARLGADQIDAVEAHGTGTRLGDPIEAQALIAAYGADRTPDRPLWLGSLKSNIGHAQAA 386

Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDW-NTPRRAGVS 423
Db        387 AGVGGLIKMLLAMRAGTLPPTLHADVPTPLVDWSAGVVRLSTGVVPWPALPGAPRRAGIS 446

Qy        424 SFGMSGTNAHVVLEEAPAATCTPPAPERPAELL-----VLSARTASALDAQAARLRDHLE 478
Db        447 AFGVSGTNAHVIVEQPPPVPVDDPAP--PTRTLPLVPWVLSGRTEAALRAQADRLRTHLA 504

Qy        479 TYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSR 538
Db        505 AHPDADPLDVGFSLATSRAALEHRAVLVAADRDGLLRLVDALAAGEPAAGLIRGTVRHDR 564

Qy        539 GKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDA 598
Db        565 -RTGFLFAGQGGQRVGMARELYEAFPAFADALDQLAARLDRHLDRPLLRVLFAEPGSDDA 623

Qy        599 ALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLV 658
Db        624 RLLDGTRYAQAALFAVEVALFRLVHGWGVRPDVLLGHSVGELAAAHVAGVLDVDDACELV 683

Qy        659 AARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAI 718
Db        684 AARGRLMGELPSGGAMVAVRATEEEVGPLL--DGQRVAVAAVNGPRSVVVSGDEEAVLAV 741

Qy        719 AAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACT-D 777
Db        742 AARCAALGHRTRRLNVSHAFHSPHVEAMLEPFRRVARGLTYHAPTIPVVSNATGRLATAD 801

Qy        778 EVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACM-PDARPALLA 836
Db        802 ALRDPGYWVRHVRQPVRFRDGVRAARDQGATAFVGLGPDGVLCALAEECLGPTGDVLLLP 861

Qy        837 SSRAGRDEPATVLEALGGLWAVGGLVSWAGLFP-SGGRRVPLPTYPWQRERYWIDTKADD 895
Db        862 VLRPGRPEPATLLAALAGAYAGGAEMDWSRVFAGTGARRVELPTYAFQHRRYWLAPGPPS 921

Qy        896 AARGDR----------RAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAG 944
Db        922 ARRDDAWRYRIAWRPLPTVPAAAGTETVAG---------AWLLVVPAHDGVASLADAAE 971

Qy        945 DRPFRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGK 1004
Db        972 ------------------RAVHR--GGATVTRLTVDAADVDRDTLAAVLTEAAADADGG 1010

Qy       1005 PNPPLLLGG---ECAGRIVAVGEGVNG-LVVGQPVIALSAGAFATHVTTSAALVLP--RP 1058
Db       1011 PDGVLCLLGLDDRAHPRSASVPRGVLATLSLAQALTDLGASARLWCVTRGAVAVTPGESP 1070

Qy       1059 QALSA-------IEAAAMPVAYLTAWYAL----------------------DRIA---- 1084
Db       1071 SVAGAQLWGFGRVAALELP----RSWGGLVDLPVDPDDRDWDLLRRALRGPEDQVAVRGA 1126

Qy       1085 -----RLQP-----GER-------VLIHAATGGVGLAAVQWAQHVGAEVHA--------- 1118
Db       1127 VGYARRLVPAPAPRAERAWRPRGTVLVTGGTGALGAHTARWLARNGA-THLVLTSRRGGN 1185

Qy       1119 TAGTPEKRAYLESLG--VRYVSDSRSDR----------FVADVRAWTGGEGVDVV---- 1161
Db       1186 APGVAALRAELVTLGAEVTVVACDVADREAVAGLLAGIPRAAPLTAVFHAAGVPQVTPLH 1245

Qy       1162 -------------------LNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGL-- 1199
Db       1246 ETTPELFAQVCAGKVAGAVHLHELAGDL--DAFVTFASAAGVWGSGGQCAYAAANAALDA 1303

Qy       1200 ------------------------------RPFLRNLSFSLVDLRGMMLERPARVRALLE 1229
Db       1304 LAERRRAAGLPATSVAWGVWGGPGMGAGAGEEYLRR--------RGVRAMPPAAALAALG 1355

Qy       1230 ELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHAGAG 1289
Db       1356 RILDADETGV----TVSDTEWGRFASGFAAARPAPLLAELPGGDVDP----------AG 1400

Qy       1290 PSTGDRDLLDRLASAAPAARAAALEAFLRTQVSQVL-RTPEIKVGAEALFTRLGMDSLMA 1348
Db       1401 PAHRAQPPVPRPAPA--ATDRPGLLALVRAEAAGVLGHDGADDVPADAEFSALGFDSLAA 1458

Qy       1349 VELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATAL 1389
Db       1459 VQLRRRLAEATGLSLSAP--------VLFDHRTPDALAAHL 1491
```

RESULT 12
US-09-808-880-3
; Sequence 3, Application US/09808880

```
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3816
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-808-880-3

  Query Match             29.5%;  Score 2128;  DB 11;  Length 3816;
  Best Local Similarity   28.0%;  Pred. No. 6.3e-154;
  Matches  617;  Conservative  216;  Mismatches  544;  Indels  828;  Gaps   44;

Qy          2 ADRPIERAA---EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAW 57
Db       1581 ADLSVVRNADEEDDPVVIVGMACRFPGGIDTPEAFWKLLEAGGDVISELPANRGWDMERL 1640

Qy         58 FDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAI 117
Db       1641 LNPDPEAKGTSATRYGGFLYDAGEFDAAFFGISPREALAMDPQQRLLLETVWELIESAGV 1700

Qy        118 APSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGP 177
Db       1701 APDSLHRSRTGTFIGSNGQFYAPLL--WNSGGDLEGYQGVGNAGSVMSGRVAYSLGLEGP 1758

Qy        178 CVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCK 237
Db       1759 AVTVDTACSSSLVALHLAVQALRRGECSLAIAGGVTVMSTPDSFVEFSRQQGLSEDGRCK 1818

Qy        238 AFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQE 297
Db       1819 AFASTADGFGLAEGVSALLVERLSDARRLGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQ 1878

Qy        298 IVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTN 357
Db       1879 RVIRAALADAGLAPADVDVVEAHGTGTRLGDPIEAQALLATYGQGRAGGRPVVLGSVKSN 1938

Qy        358 LGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISW--GDLRLTVTRARTPWPDWN 415
Db       1939 IGHTQAAAGVAGVMKMVLALERGVVPKTLHVDEPSPHVDWSAGEVELAV--EAVPWSRGG 1996

Qy        416 TPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELL---------VLSARTASAL 466
Db       1997 RVRRAGVSSFGISGTNAHVIVEEAPA----EPEPEPGTRVVAAGDLVVPWVVSGRDAGAL 2052

Qy        467 DAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTS 526
Db       2053 REQAARLAAHVSSTGAGVV-DVGWSLVATRSVFEHRAVMVGSELDSMAESLAGFAAGGVV 2111

Qy        527 PGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLR 586
Db       2112 PGVVSGVAPAEGRRVVFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLV 2171

Qy        587 EVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVA 646
Db       2172 EVL----RGGGEAVLGRVDVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVA 2227

Qy        647 GVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQV 706
Db       2228 GGLSLADGARVVVLRSRAIARIAGGGGMVSVGLSAERVRTMLDTYGGRVSVAAVNGPSST 2287

Qy        707 VIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVL 766
Db       2288 VVSGDVQALDELLAGCEREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQHSSVPF 2347

Qy        767 VSNLSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPAC 826
Db       2348 FSTVTADWLDTTALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQT 2407

Qy        827 MP--DARPALLASSRAGRDEPA--TVLEALGGLWAVGGLVSWA----GLFPSGGRRVPLP 878
Db       2408 LDALDQNAAVLGSLR--RDEGGLDRLLTSLAEAFVQGVPVDWTHAFEGMTP---RTVDLP 2462

Qy        879 TYPWQRERYW-------------------------------------------------- 888
Db       2463 TYPFQRQHYWPKPAPAPGANLGDVASVGLTAAGHPLLGAVVEMPDSDGLVLTGQISLRTH 2522

Qy        889 ------------------------------------------------------------ 888
Db       2523 PWLADHEVLGSVLLPGTAFVELAVQAADRAGYDVLDELTLEAPLVLPDRGGIQVRLALGP 2582

Qy        889 ----------IDTKADDAA---RGDRRAPG------------------------------ 905
Db       2583 SEADGRRSLQLHSRPEEAAGFHRWTRHASGFVVPGGTGAARPTEPAGVWPPAGAEPVALA 2642

Qy        906 ------------------------------------------------------------ 905
Db       2643 SDRYARLVERGYTYGPSFQGLHTAWRHGDDVYAEVALPEGTPADGYALHPALLDAAVQAV 2702

Qy        906 --------------------------------------AGHDE----------------- 910
Db       2703 GLGSFVEDPGQVYLPFLWSDVTLHATGATSLRVRVSPAGPDTVALALADPAGAPVATVGA 2762

Qy        911 ---------------------------VEEGGAV---RGG-------DRRSARL------ 927
Db       2763 LRLRTTSAAQLARARGSAEHAMFRVEWVEEGSAADRCRGGAGGTTYEGERAAEAGAAAGT 2822

Qy        928 --------------------------DHP-PPES-------------------------- 934
Db       2823 WAVLGPRVPAAVRTMGVDVVTALDTPDHPADPQSLADLAALGDTVPDVVVVTSLLSLASG 2882

Qy        935 -----GRREKVEAA---------------------------------------------- 943
Db       2883 ADSPLGNRPRPTAAEQDTAATVAGVHSALHAALDLVQAWLADERHTASRLVLVTRHAMTV 2942

Qy        944 -------------------------------------GDR-------------------- 946
Db       2943 AESDPEPDLLLAPVWGLVRSAQAENPGRFVLADIDGDEASWDALPRAVASAASEVAIRAG 3002

Qy        947 -----------------------PFRLEIDEPGVLDHLVLRVTERRAPGLG--EVEIAVD 981
Db       3003 AVYVPRLARATDEGLVVADEAAGPWRLDVTEAGTLANLALVPCPDASRPLGPDEVRIAVR 3062

Qy        982 AAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAG 1041
Db       3063 AAGVNFRDVLLALGMYPDE---------GLMGAEAAGVVTEVGGGVTTLAPGDRVMGLVTG 3114

Qy       1042 AFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGV 1101
Db       3115 GFGPVAVTHHRMLVRMPRGWSFAEAASVPVAFLTAYYALHDLAGLRGGESVLVHSAAGGV 3174

Qy       1102 GLAAVQWAQHVGAEVHATAGTPEKRAYLESLGV--RYVSDSRSDRFVADVRAWTGGEGVD 1159
Db       3175 GMAAVQLARHWDAEVFGTA-SKGKWDVLAAQGLDEEHIGSSRTTEFEQRFRATSGGRGID 3233

Qy       1160 VVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGL-----RPFLRNLSFSLVDLR 1214
Db       3234 VVLNALSGDFVDASARLLREGGRFVEMGKTDIRTD--LGVVGADGVPDIRYVAFDLAEAG 3291
```

```
Qy       1215 GMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLT-- 1272
Db       3292 A------ERIGQMLDEIMALFDAGVLRLPPLRAWPVRRAHEALRFVSQARHVGKVVLTVP 3345

Qy       1273 --------------------------------------------------LGDPEVQIRI- 1282
Db       3346 AALDAEGTVLITGAGTLGALVARHLVTEHDVRRLLLVSRSGVAPDLAAELGALGAEVTVA 3405

Qy       1283 ------------------PTH--------AG---------------------------- 1287
Db       3406 ACDVANRKALKALLEDIPPEHPVTGIVHTAGVLDDGVVSGLTPERVDTVLKPKVDAALTL 3465

Qy       1288 ------------------------AGPSTG---------------------------- 1293
Db       3466 ESVIGELDLDPALFVIFSSAASMLGGPGQGSYAAANQFLDTLARHRARRGLTSVSLGWGL 3525

Qy       1294 --------------DRD-----------------LLDRL--------------------- 1301
Db       3526 WHEASGLTGGLADIDRDRMSRAGIAPMPTDEALHLFDRATELGDPVLLPMRLNEAALEDR 3585

Qy       1302 -----------------------ASAAPAARAAALEAF-----------------LRTQV 1321
Db       3586 AADGTLPPLLSGLVRVRHRPSARAGTATAAPATGPEAFARELAAAPDPRRALRDLVRGHV 3645

Qy       1322 SQVL--RTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQ 1379
Db       3646 ALVLGHSGPE-AIDAEQAFRDIGFDSLTAVELRNRLNAETGLRLPGTLVFDYPNPSALAD 3704

Qy       1380 NLLDALATA---------LSLERVAAENLRAGVQNDFVSSGADQD 1415
Db       3705 HLLELLAPATQPTAAPLLAELERV-EQLLSAAASPGGPASAVDEE 3748
```

RESULT 13
US-09-861-289-2
; Sequence 2, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 5215
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-2

```
  Query Match             29.5%;  Score 2125;  DB 10;  Length 5215;
  Best Local Similarity   37.0%;  Pred. No. 1.7e-153;
  Matches  581;  Conservative  175;  Mismatches  568;  Indels  248;  Gaps   46;

Qy          1 VADRPI---ERAA-EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAA 55
Db       2812 VAERGASFGDRATDDDPIAIVGMACRYPGGVSSPEDLWRLVAEGTDAISEFPVNRGWDLE 2871

Qy         56 AWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENA 115
Db       2872 SLYDPDPESKGTTYCREGGFLEGAGDFDAAFFGISPREALVMDPQQRLLLEVSWEALERA 2931

Qy        116 AIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGL--GTMPSVGAGRISYALG 173
Db       2932 GIDPSSLRGSRGGVYVGAAHGSY-ASDPRLVPEG---SEGYLLTGSADAVMSGRISYALG 2987

Qy        174 LRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARD 233
Db       2988 LEGPSMTVETACSSSLVALHLAVRALRHGECGLALAGGVAVMADPAAFVEFSRQKGLAAD 3047

Qy        234 GRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNG 293
Db       3048 GRCKAFSAAADGTGWAEGVGVLVLERLSDARRAGHTVLGLVTGTAVNQDGASNGLTAPNG 3107

Qy        294 SSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGS 353
Db       3108 PAQQRVIAEALADAGLSPEDVDAVEAHGTGTRLGDPIEAGALLAASGRNRSGDHPLWLGS 3167

Qy        354 VKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPD 413
Db       3168 LKSNIGHAQAAAGVGGVIKMLQALRHGLLPRTLHADEPTPHADWSSGRVRLLTSEVPWQR 3227

Qy        414 WNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE---------RPAE------LLVL 458
Db       3228 TGRPRRTGVSAFGVGGTNAHVVLEEAPA----PPAPEPAGEAPGGSRAAEGAEGPLAWVV 3283

Qy        459 SARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALD 518
Db       3284 SGRDEPALRSQARRLRDHLSRTPGARPRDIAFSLAATRAAFDHRAVLIGSDGAELAAALD 3343

Qy        519 AAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFN 578
Db       3344 ALAEGRDGPAVVRGVRDRD-GRMAFLFTGQGSQRAGMAHDLHAAHTFFASALDEVTDRLD 3402

Qy        579 QELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIG 638
Db       3403 PLLGRPLGALLDARPGSPEAALLDRTEYTQPALFAVEVALHRLLEHWGMRPDLLLGHSVG 3462

Qy        639 ELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIA 698
Db       3463 ELAAAHVAGVLDLDDACALVAARGRLMQRLPPGGAMVSVRAGEDEVRALLAGREDAVCVA 3522

Qy        699 AVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVS 758
Db       3523 AVNGPRSVVISGAEEAVAEAAAQLAGRGRRTRRLRVAHAFHSPLMDGMLAGFREVAAGLR 3582

Qy        759 YRRPSIVLVSNLSGK-ACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKS 817
Db       3583 YREPELTVVSTVTGRPARPGELTGPDYWVAQVREPVRFADAVRTAHRLGARTFLETGPDG 3642

Qy        818 TLLGLVPACMPDARPALLASSRAGRDEPAT---------VLEALGGLWAVGGLVSWAGLF 868
Db       3643 VLCGMAEECLEDDTVALLPAIH----KPGTAPHGPAAPGALRAAAAAYGRGARVDWAGMH 3698

Qy        869 PSG----GRRVPLPTYPWQRERYWIDTKADDAARGDRRAPG-----------AGHDEVEE 913
Db       3699 ADGPEGPARRVELPVHAFRHRRYWL-------------APGRAADTDDWMYRIGWDRLP- 3744

Qy        914 GGAVRGGDRRSAR--LDHPPPE-----SGRREK-VEAAGDRPFRLEIDEPGVLDHLVLRV 965
Db       3745 --AVTGGARTAGRWLVIHPDSPRCRELSGHAERALRAAGASPVPLPVDAPAA-DRASFAA 3801

Qy        966 TERRAPGLG-EVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAV-- 1022
Db       3802 LLRSATGPDTRGDTAAPVAGV------LSL-LSEEDRPHRQHAPVPAG------VLATLS 3848

Qy       1023 ------GEGVNGLV--VGQPVIA-------LSAGAFATHVTTSAALVLP----------- 1056
Db       3849 LMQAMEEEAVEARVWCVSRAAVAAADRERPVGAGAALWGLGRVAALERPTRWGGLVDLPA 3908

Qy       1057 RPQA---LSAIEAAAMP-----VAYLTAW-YALDRIAR----------LQPGERVLIHAA 1097
Db       3909 SPGAAHWAAAVERLAGPEDQIAVRASGSWGRRLTRLPRDGGGRTAAPAYRPRGTVLVTGG 3968

Qy       1098 TGGVGLAAVQWAQHVGAE----------------------------VHATAGTPEK--- 1125
Db       3969 TGALGGHLARWLAAAGAEHLALTSRRGPDAPGAAGLEAELLLLGAKVTFAACDTADRDGL 4028

Qy       1126 -------------RAYLESLGVRYVS--DSRSDRFVADVRAW--TGGEGVDVVLNSLSGE 1168
```

```
Db     4029 ARVLRAIPEDTPLTAVFHAAGVPQVTPLSRTSPEHFADVYAGKAAGAAHLDELTRELGAG 4088

Qy     1169 LIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR------------------------ 1204

Db     4089 L--DAFVLYSSGAGVWGSAGQGAYAAANAALDALARRRAADGLPATSIAWGVWGGGGMGA 4146

Qy     1205 -NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIAT-LPIARVADAFRSMAQ 1262

Db     4147 DEAGAEYLGRRGM---RPMAPVSALRAMATAIASG--EPCPTVTHTDWERFGEGFTAFRP 4201

Qy     1263 AQHLGKLVLTLGDPEVQIRIPTHAGAG-----PSTGDRDLLDRLASAAPAARAAALEAFL 1317

Db     4202 S----PLIAGLGTP----------GGGRAAETPEEGNATAAADLTALPPAELRTALRELV 4247

Qy     1318 RTQVSQVLRTPE-IKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIAL 1376

Db     4248 RARTAAALGLDDPAEVAEGERFPAMGFDSLATVRLRRGLASATGLDLPPDLLFDRDTPAA 4307

Qy     1377 LAQNLLDALATA 1388

Db     4308 LAAHLAELLATA 4319
```

RESULT 14
US-09-860-846-2
; Sequence 2, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-2

```
  Query Match             29.5%;  Score 2125;  DB 10;  Length 5215;
  Best Local Similarity   37.0%;  Pred. No. 1.7e-153;
  Matches  581;  Conservative  175;  Mismatches  568;  Indels  248;  Gaps   46;

Qy        1 VADRPI---ERAA-EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAA 55

Db     2812 VAERGASFGDRATDDDPIAIVGMACRYPGGVSSPEDLWRLVAEGTDAISEFPVNRGWDLE 2871

Qy       56 AWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENA 115

Db     2872 SLYDPDPESKGTTYCREGGFLEGAGDFDAAFFGISPREALVMDPQQRLLLEVSWEALERA 2931

Qy      116 AIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGL--GTMPSVGAGRISYALG 173

Db     2932 GIDPSSLRGSRGGVYVGAAHGSY-ASDPRLVPEG---SEGYLLTGSADAVMSGRISYALG 2987

Qy      174 LRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARD 233

Db     2988 LEGPSMTVETACSSSLVALHLAVRALRHGECGLALAGGVAVMADPAAFVEFSRQKGLAAD 3047

Qy      234 GRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNG 293

Db     3048 GRCKAFSAAADGTGWAEGVGVLVLERLSDARRAGHTVLGLVTGTAVNQDGASNGLTAPNG 3107

Qy      294 SSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGS 353

Db     3108 PAQQRVIAEALADAGLSPEDVDAVEAHGTGTRLGDPIEAGALLAASGRNRSGDHPLWLGS 3167

Qy      354 VKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPD 413

Db     3168 LKSNIGHAQAAAGVGGVIKMLQALRHGLLPRTLHADEPTPHADWSSGRVRLLTSEVPWQR 3227

Qy      414 WNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE---------RPAE------LLVL 458

Db     3228 TGRPRRTGVSAFGVGGTNAHVVLEEAPA----PPAPEPAGEAPGGSRAAEGAEGPLAWVV 3283

Qy      459 SARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALD 518

Db     3284 SGRDEPALRSQARRLRDHLSRTPGARPRDIAFSLAATRAAFDHRAVLIGSDGAELAAALD 3343

Qy      519 AAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFN 578

Db     3344 ALAEGRDGPAVVRGVRDRD-GRMAFLFTGQGSQRAGMAHDLHAAHTFFASALDEVTDRLD 3402

Qy      579 QELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIG 638

Db     3403 PLLGRPLGALLDARPGSPEAALLDRTEYTQPALFAVEVALHRLLEHWGMRPDLLLGHSVG 3462

Qy      639 ELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIA 698

Db     3463 ELAAAHVAGVLDLDDACALVAARGRLMQRLPPGGAMVSVRAGEDEVRALLAGREDAVCVA 3522

Qy      699 AVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVS 758

Db     3523 AVNGPRSVVISGAEEAVAEAAAQLAGRGRRTRRLRVAHAFHSPLMDGMLAGFREVAAGLR 3582

Qy      759 YRRPSIVLVSNLSGK-ACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKS 817

Db     3583 YREPELTVVSTVTGRPARPGELTGPDYWVAQVREPVRFADAVRTAHRLGARTFLETGPDG 3642

Qy      818 TLLGLVPACMPDARPALLASSRAGRDEPAT---------VLEALGGLWAVGGLVSWAGLF 868

Db     3643 VLCGMAEECLEDDTVALLPAIH----KPGTAPHGPAAPGALRAAAAAYGRGARVDWAGMH 3698

Qy      869 PSG----GRRVPLPTYPWQRERYWIDTKADDAARGDRRAPG-----------AGHDEVEE 913

Db     3699 ADGPEGPARRVELPVHAFRHRRYWL-------------APGRAADTDDWMYRIGWDRLP- 3744

Qy      914 GGAVRGGDRRSAR--LDHPPPE-----SGRREK-VEAAGDRPFRLEIDEPGVLDHLVLRV 965

Db     3745 --AVTGGARTAGRWLVIHPDSPRCRELSGHAERALRAAGASPVPLPVDAPAA-DRASFAA 3801

Qy      966 TERRAPGLG-EVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAV-- 1022

Db     3802 LLRSATGPDTRGDTAAPVAGV------LSL-LSEEDRPHRQHAPVPAG------VLATLS 3848

Qy     1023 ------GEGVNGLV--VGQPVIA-------LSAGAFATHVTTSAALVLP----------- 1056

Db     3849 LMQAMEEEAVEARVWCVSRAAVAAADRERPVGAGAALWGLGRVAALERPTRWGGLVDLPA 3908

Qy     1057 RPQA---LSAIEAAAMP-----VAYLTAW-YALDRIAR----------LQPGERVLIHAA 1097

Db     3909 SPGAAHWAAAVERLAGPEDQIAVRASGSWGRRLTRLPRDGGGRTAAPAYRPRGTVLVTGG 3968

Qy     1098 TGGVGLAAVQWAQHVGAE-----------------------------VHATAGTPEK--- 1125

Db     3969 TGALGGHLARWLAAAGAEHLALTSRRGPDAPGAAGLEAELLLGAKVTFAACDTADRDGL 4028

Qy     1126 -------------RAYLESLGVRYVS--DSRSDRFVADVRAW--TGGEGVDVVLNSLSGE 1168

Db     4029 ARVLRAIPEDTPLTAVFHAAGVPQVTPLSRTSPEHFADVYAGKAAGAAHLDELTRELGAG 4088

Qy     1169 LIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR------------------------ 1204

Db     4089 L--DAFVLYSSGAGVWGSAGQGAYAAANAALDALARRRAADGLPATSIAWGVWGGGGMGA 4146

Qy     1205 -NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIAT-LPIARVADAFRSMAQ 1262

Db     4147 DEAGAEYLGRRGM---RPMAPVSALRAMATAIASG--EPCPTVTHTDWERFGEGFTAFRP 4201
```

```
Qy      1263 AQHLGKLVLTLGDPEVQIRIPTHAGAG-----PSTGDRDLLDRLASAAPAARAAALEAFL 1317
Db      4202 S----PLIAGLGTP----------GGGRAAETPEEGNATAAADLTALPPAELRTALRELV 4247

Qy      1318 RTQVSQVLRTPE-IKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIAL 1376
Db      4248 RARTAAALGLDDPAEVAEGERFPAMGFDSLATVRLRRGLASATGLDLPPDLLFDRDTPAA 4307

Qy      1377 LAQNLLDALATA 1388
Db      4308 LAAHLAELLATA 4319
```

RESULT 15
US-09-988-384B-2
; Sequence 2, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin.
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-2

```
  Query Match             29.5%;  Score 2125;  DB 11;  Length 5215;
  Best Local Similarity   37.0%;  Pred. No. 1.7e-153;
  Matches  581;  Conservative  175;  Mismatches  568;  Indels  248;  Gaps   46;

Qy         1 VADRPI---ERAA-EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAA 55
Db      2812 VAERGASFGDRATDDDPIAIVGMACRYPGGVSSPEDLWRLVAEGTDAISEFPVNRGWDLE 2871

Qy        56 AWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENA 115
Db      2872 SLYDPDPESKGTTYCREGGFLEGAGDFDAAFFGISPREALVMDPQQRLLLEVSWEALERA 2931

Qy       116 AIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGL--GTMPSVGAGRISYALG 173
Db      2932 GIDPSSLRGSRGGVYVGAAHGSY-ASDPRLVPEG---SEGYLLTGSADAVMSGRISYALG 2987

Qy       174 LRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARD 233
Db      2988 LEGPSMTVETACSSSLVALHLAVRALRHGECGLALAGGVAVMADPAAFVEFSRQKGLAAD 3047

Qy       234 GRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNG 293
Db      3048 GRCKAFSAAADGTGWAEGVGVLVLERLSDARRAGHTVLGLVTGTAVNQDGASNGLTAPNG 3107

Qy       294 SSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGS 353
Db      3108 PAQQRVIAEALADAGLSPEDVDAVEAHGTGTRLGDPIEAGALLAASGRNRSGDHPLWLGS 3167

Qy       354 VKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPD 413
Db      3168 LKSNIGHAQAAAGVGGVIKMLQALRHGLLPRTLHADEPTPHADWSSGRVRLLTSEVPWQR 3227

Qy       414 WNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE---------RPAE------LLVL 458
Db      3228 TGRPRRTGVSAFGVGGTNAHVVLEEAPA----PPAPEPAGEAPGGSRAAEGAEGPLAWVV 3283

Qy       459 SARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALD 518
Db      3284 SGRDEPALRSQARRLRDHLSRTPGARPRDIAFSLAATRAAFDHRAVLIGSDGAELAAALD 3343

Qy       519 AAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFN 578
Db      3344 ALAEGRDGPAVVRGVRDRD-GRMAFLFTGQGSQRAGMAHDLHAAHTFFASALDEVTDRLD 3402

Qy       579 QELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIG 638
Db      3403 PLLGRPLGALLDARPGSPEAALLDRTEYTQPALFAVEVALHRLLEHWGMRPDLLLGHSVG 3462

Qy       639 ELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIA 698
Db      3463 ELAAAHVAGVLDLDDACALVAARGRLMQRLPPGGAMVSVRAGEDEVRALLAGREDAVCVA 3522

Qy       699 AVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVS 758
Db      3523 AVNGPRSVVISGAEEAVAEAAAQLAGRGRRTRRLRVAHAFHSPLMDGMLAGFREVAAGLR 3582

Qy       759 YRRPSIVLVSNLSGK-ACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKS 817
Db      3583 YREPELTVVSTVTGRPARPGELTGPDYWVAQVREPVRFADAVRTAHRLGARTFLETGPDG 3642

Qy       818 TLLGLVPACMPDARPALLASSRAGRDEPAT---------VLEALGGLWAVGGLVSWAGLF 868
Db      3643 VLCGMAEECLEDDTVALLPAIH----KPGTAPHGPAAPGALRAAAAAYGRGARVDWAGMH 3698

Qy       869 PSG----GRRVPLPTYPWQRERYWIDTKADDAARGDRRAPG-----------AGHDEVEE 913
Db      3699 ADGPEGPARRVELPVHAFRHRRYWL-------------APGRAADTDDWMYRIGWDRLP- 3744

Qy       914 GGAVRGGDRRSAR--LDHPPPE-----SGRREK-VEAAGDRPFRLEIDEPGVLDHLVLRV 965
Db      3745 --AVTGGARTAGRWLVIHPDSPRCRELSGHAERALRAAGASPVPLPVDAPAA-DRASFAA 3801

Qy       966 TERRAPGLG-EVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAV-- 1022
Db      3802 LLRSATGPDTRGDTAAPVAGV------LSL-LSEEDRPHRQHAPVPAG------VLATLS 3848

Qy      1023 ------GEGVNGLV--VGQPVIA-------LSAGAFATHVTTSAALVLP----------- 1056
Db      3849 LMQAMEEEAVEARVWCVSRAAVAAADRERPVGAGAALWGLGRVAALERPTRWGGLVDLPA 3908

Qy      1057 RPQA---LSAIEAAMP-----VAYLTAW-YALDRIAR----------LQPGERVLIHAA 1097
Db      3909 SPGAAHWAAAVERLAGPEDQIAVRASGSWGRRLTRLPRDGGGRTAAPAYRPRGTVLVTGG 3968

Qy      1098 TGGVGLAAVQWAQHVGAE----------------------------VHATAGTPEK--- 1125
Db      3969 TGALGGHLARWLAAAGAEHLALTSRRGPDAPGAAGLEAELLLLGAKVTFAACDTADRDGL 4028

Qy      1126 -------------RAYLESLGVRYVS--DSRSDRFVADVRAW--TGGEGVDVVLNSLSGE 1168
Db      4029 ARVLRAIPEDTPLTAVFHAAGVPQVTPLSRTSPEHFADVYAGKAAGAAHLDELTRELGAG 4088

Qy      1169 LIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR------------------------ 1204
Db      4089 L--DAFVLYSSGAGVWGSAGQGAYAAANAALDALARRRAADGLPATSIAWGVWGGGGMGA 4146

Qy      1205 -NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIAT-LPIARVADAFRSMAQ 1262
Db      4147 DEAGAEYLGRRGM---RPMAPVSALRAMATAIASG--EPCPTVTHTDWERFGEGFTAFRP 4201

Qy      1263 AQHLGKLVLTLGDPEVQIRIPTHAGAG-----PSTGDRDLLDRLASAAPAARAAALEAFL 1317
Db      4202 S----PLIAGLGTP----------GGGRAAETPEEGNATAAADLTALPPAELRTALRELV 4247

Qy      1318 RTQVSQVLRTPE-IKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIAL 1376
Db      4248 RARTAAALGLDDPAEVAEGERFPAMGFDSLATVRLRRGLASATGLDLPPDLLFDRDTPAA 4307
```

```
Qy       1377 LAQNLLDALATA 1388
              ||  :|  : ||||
Db       4308 LAAHLAELLATA 4319
```

Search completed: October 2, 2003, 17:47:41
Job time : 109.09 secs

```
OM protein - protein search, using sw model

Run on:         October  2, 2003, 17:27:33 ; Search time 34.7527 Seconds
                                             (without alignments)
                                             1730.044 Million cell updates/sec

Title:          US-10-014-717-2
Perfect score:  7210
Sequence:       1 VADRPIERAAEDPIAIVGAS..........GVQNDFVSSGADQDWEIIAL 1421

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters:      328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                 1:  /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
                 2:  /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
                 3:  /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
                 4:  /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
                 5:  /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                 6:  /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7210	100.0	1421	3	US-09-335-409-2	Sequence 2, Appli
2	7210	100.0	1421	4	US-09-568-102-2	Sequence 2, Appli
3	7210	100.0	1421	4	US-09-567-969-2	Sequence 2, Appli
4	7210	100.0	1421	4	US-09-568-480-2	Sequence 2, Appli
5	7210	100.0	1421	4	US-09-568-486-2	Sequence 2, Appli
6	7210	100.0	1421	4	US-09-568-472-2	Sequence 2, Appli
7	7210	100.0	1421	4	US-09-567-899-2	Sequence 2, Appli
8	4137	57.4	2439	3	US-09-335-409-7	Sequence 7, Appli
9	4137	57.4	2439	4	US-09-568-102-7	Sequence 7, Appli
10	4137	57.4	2439	4	US-09-567-969-7	Sequence 7, Appli
11	4137	57.4	2439	4	US-09-568-480-7	Sequence 7, Appli
12	4137	57.4	2439	4	US-09-568-486-7	Sequence 7, Appli
13	4137	57.4	2439	4	US-09-568-472-7	Sequence 7, Appli
14	4137	57.4	2439	4	US-09-567-899-7	Sequence 7, Appli
15	3831.5	53.1	2259	3	US-09-413-814-70	Sequence 70, Appl
16	3311.5	45.9	7257	3	US-09-335-409-5	Sequence 5, Appli
17	3311.5	45.9	7257	4	US-09-568-102-5	Sequence 5, Appli
18	3311.5	45.9	7257	4	US-09-567-969-5	Sequence 5, Appli
19	3311.5	45.9	7257	4	US-09-568-480-5	Sequence 5, Appli
20	3311.5	45.9	7257	4	US-09-568-486-5	Sequence 5, Appli
21	3311.5	45.9	7257	4	US-09-568-472-5	Sequence 5, Appli
22	3311.5	45.9	7257	4	US-09-567-899-5	Sequence 5, Appli
23	2273	31.5	2152	3	US-09-036-987A-3	Sequence 3, Appli
24	2273	31.5	2152	3	US-09-370-700-3	Sequence 3, Appli
25	2273	31.5	2152	4	US-09-603-207-3	Sequence 3, Appli
26	2250	31.2	6095	3	US-09-144-085-2	Sequence 2, Appli
27	2245.5	31.1	2890	3	US-09-413-814-67	Sequence 67, Appl
28	2245.5	31.1	3798	3	US-09-335-409-6	Sequence 6, Appli
29	2245.5	31.1	3798	4	US-09-568-102-6	Sequence 6, Appli
30	2245.5	31.1	3798	4	US-09-567-969-6	Sequence 6, Appli
31	2245.5	31.1	3798	4	US-09-568-480-6	Sequence 6, Appli
32	2245.5	31.1	3798	4	US-09-568-486-6	Sequence 6, Appli
33	2245.5	31.1	3798	4	US-09-568-472-6	Sequence 6, Appli
34	2245.5	31.1	3798	4	US-09-567-899-6	Sequence 6, Appli
35	2194	30.4	5087	3	US-09-144-085-1	Sequence 1, Appli
36	2150	29.8	1841	2	US-08-804-227C-6	Sequence 6, Appli
37	2128	29.5	3816	3	US-09-428-517-3	Sequence 3, Appli
38	2127	29.5	3170	3	US-09-036-987A-4	Sequence 4, Appli
39	2127	29.5	3170	3	US-09-370-700-4	Sequence 4, Appli
40	2127	29.5	3170	4	US-09-603-207-4	Sequence 4, Appli
41	2125	29.5	5215	3	US-09-105-537-2	Sequence 2, Appli
42	2113	29.3	4928	3	US-09-036-987A-5	Sequence 5, Appli
43	2113	29.3	4928	3	US-09-370-700-5	Sequence 5, Appli
44	2113	29.3	4928	4	US-09-603-207-5	Sequence 5, Appli
45	2108.5	29.2	3562	4	US-09-679-279-14	Sequence 14, Appl

ALIGNMENTS

```
RESULT 1
US-09-335-409-2
; Sequence 2, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1421
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-335-409-2

  Query Match             100.0%;  Score 7210;  DB 3;  Length 1421;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1421;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60

Qy         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120

Qy        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180

Qy        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240

Qy        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Qy        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
```

```
Qy        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Qy        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480

Qy        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660

Qy        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720

Qy        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780

Qy        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840

Qy        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900

Qy        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960

Qy        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020

Qy       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080

Qy       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140

Qy       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200

Qy       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260

Qy       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320

Qy       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380

Qy       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
              |||||||||||||||||||||||||||||||||||||||||
Db       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
```

```
RESULT 2
US-09-568-102-2
; Sequence 2, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,102
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1421
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-568-102-2

  Query Match             100.0%;  Score 7210;  DB 4;  Length 1421;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1421;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60

Qy         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120

Qy        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180

Qy        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240

Qy        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Qy        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360

Qy        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Qy        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480

Qy        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
```

```
Qy     661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720

Qy     721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780

Qy     781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840

Qy     841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900

Qy     901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960

Qy     961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020

Qy    1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080

Qy    1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140

Qy    1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200

Qy    1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260

Qy    1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320

Qy    1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380

Qy    1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
           |||||||||||||||||||||||||||||||||||||||||
Db    1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
```

```
RESULT 3
US-09-567-969-2
; Sequence 2, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1421
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-567-969-2

  Query Match             100.0%;  Score 7210;  DB 4;  Length 1421;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1421;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60

Qy      61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120

Qy     121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180

Qy     181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240

Qy     241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Qy     301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360

Qy     361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Qy     421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480

Qy     481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540

Qy     541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Qy     601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660

Qy     661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720

Qy     721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780

Qy     781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840

Qy     841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900

Qy     901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
```

```
Qy        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020

Qy       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080

Qy       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140

Qy       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200

Qy       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260

Qy       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320

Qy       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380

Qy       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
              |||||||||||||||||||||||||||||||||||||||||
Db       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
```

```
RESULT 4
US-09-568-480-2
; Sequence 2, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1421
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-568-480-2

  Query Match             100.0%;  Score 7210;  DB 4;  Length 1421;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1421;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60

Qy         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120

Qy        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180

Qy        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240

Qy        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Qy        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360

Qy        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Qy        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480

Qy        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660

Qy        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720

Qy        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780

Qy        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840

Qy        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900

Qy        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960

Qy        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020

Qy       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080

Qy       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140

Qy       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200

Qy       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
```

```
Qy       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320

Qy       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380

Qy       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
              |||||||||||||||||||||||||||||||||||||||||
Db       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
```

RESULT 5
US-09-568-486-2
; Sequence 2, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1421
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-486-2

```
  Query Match             100.0%;  Score 7210;  DB 4;  Length 1421;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1421;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60

Qy         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120

Qy        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180

Qy        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240

Qy        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Qy        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360

Qy        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Qy        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480

Qy        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660

Qy        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720

Qy        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780

Qy        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840

Qy        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900

Qy        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960

Qy        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020

Qy       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080

Qy       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140

Qy       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200

Qy       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260

Qy       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320

Qy       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380

Qy       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
              |||||||||||||||||||||||||||||||||||||||||
Db       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
```

RESULT 6
US-09-568-472-2
; Sequence 2, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas

```
; APPLICANT:  Ligon, James
; APPLICANT:  Molnar, Istvan
; APPLICANT:  Zirkle, Ross
; APPLICANT:  Cyr, Devon
; APPLICANT:  Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE:  2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1421
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-568-472-2

  Query Match             100.0%;  Score 7210;  DB 4;  Length 1421;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1421;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60

Qy         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120

Qy        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180

Qy        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240

Qy        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Qy        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360

Qy        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Qy        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480

Qy        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660

Qy        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720

Qy        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780

Qy        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840

Qy        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900

Qy        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960

Qy        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020

Qy       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080

Qy       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140

Qy       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200

Qy       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260

Qy       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320

Qy       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380

Qy       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
              |||||||||||||||||||||||||||||||||||||||||
Db       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421


RESULT 7
US-09-567-899-2
; Sequence 2, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1421
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-567-899-2

  Query Match             100.0%;  Score 7210;  DB 4;  Length 1421;
```

```
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1421;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60

Qy       61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120

Qy      121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180

Qy      181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240

Qy      241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Qy      301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360

Qy      361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Qy      421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480

Qy      481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540

Qy      541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Qy      601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660

Qy      661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720

Qy      721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780

Qy      781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840

Qy      841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900

Qy      901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960

Qy      961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020

Qy     1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080

Qy     1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140

Qy     1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200

Qy     1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260

Qy     1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320

Qy     1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380

Qy     1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
            |||||||||||||||||||||||||||||||||||||||||
Db     1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
```

RESULT 8
US-09-335-409-7
; Sequence 7, Application US/09335409.
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-7

```
  Query Match             57.4%;  Score 4137;  DB 3;  Length 2439;
  Best Local Similarity   43.4%;  Pred. No. 8.5e-297;
  Matches  933;  Conservative  135;  Mismatches  308;  Indels  776;  Gaps   13;

Qy        6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65
            :|:   :||||||  || |||      || ||:  || |   | :|  |     |  : |
Db       26 LEQERTEPIAIVGIGCRFPGGADTPEAFWELLDSGRDAV--QPLDRRWALVGVHPSEEVP 83

Qy       66 GKTPVTRASFLSD-VACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124
                   |  |:: |  |||:||| |||||  :||  |||||| || ||:| ||| :| |
Db       84 -----RWAGLLTEAVDGFDAAFFGTSPREARSLDPQQRLLLEVTWEGLEDAGIAPQSLDG 138

Qy      125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184
            : ||||:|   |:|   : |     | ||:   |   || |||:|| |||:|||: ||||
Db      139 SRTGVFLGACSSDYSHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPCLTVDTA 197

Qy      185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244
             ||||||:||||:|||: |   ||||||:::||  |:: | : :||: || |: | | |:
Db      198 CSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHCRTFDASAN 257

Qy      245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304
            || ||||| :||||||| |:  |||| |:|||||:| || |:||  ||  :|| :|: ||
Db      258 GFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLREAL 317
```

```
Qy    305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364
Db    318 QSARVDAGAIGYVETHGTGTSLGDPIEVEALRAVLGPARADGSRCVLGAVKTNLGHLEGA 377
Qy    365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424
Db    378 AGVAGLIKAALALHHELIPRNLHFHTLNPRIRIEGTALALATEPVPWPRAGRPRFAGVSA 437
Qy    425 FGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQC 484
Db    438 FGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSAAALDAQAARLSAHIAAYPEQG 497
Qy    485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544
Db    498 LGDVAFSLVSTRSPMEHRLAVAATSREALRSALEVAAQGQTPAGAARGRAASSPGKLAFL 557
Qy    545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604
Db    558 FAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCEVMWAEPGSSRSSLLDQT 617
Qy    605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664
Db    618 AFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAGVFSLEDAVRLVVARGRL 677
Qy    665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724
Db    678 MQALPAGGAMVSIAAPEADVAAAVAPHAALVSIAAVNGPEQVVIAGAEKFVQQIAAAFAA 737
Qy    725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784
Db    738 RGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSIALVSNLSGKPCTDEVSAPGY 797
Qy    785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844
Db    798 WVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACLPDARPVLLPASRAGRDE 857
Qy    845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI----DTKADDAAR-- 898
Db    858 AASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERYWIEAPVDREADGTGRAR 917
Qy    899 ------------------------------------------------------------ 898
Db    918 AGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEVVFPGAGYLEMALSSGAE 977
Qy    899 -------------------------------------------------GDRRAP----- 904
Db    978 ILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQVASREPGERRAPFRIHA 1037
Qy    905 ------------------------------------------------------------ 904
Db   1038 RGVLRRIGRVETPARSNLAALRARLHAAVPAAAIYGALAEMGLQYGPALRGLAELWRGEG 1097
Qy    905 ------------------------------------------------------------ 904
Db   1098 EALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSP 1157
Qy    905 ------------------------------------------------------------ 904
Db   1158 GELWCHARVVSDGQQASSRWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLE 1217
Qy    905 ------------------------------------GAGH-------------------- 908
Db   1218 LDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVVHAAGDDTSTAGMRALLA 1277
Qy    909 ------------------------------------------------------------ 908
Db   1278 NAFDGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDADALESALMRGCDSVLSL 1337
Qy    909 ------------------------------------------------------------ 908
Db   1338 VQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDP 1397

Qy    909 ----------------DEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEI 952
Db   1398 AEPEGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPDAQRREKVEPAGDRPFRLEI 1457
Qy    953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012
Db   1458 DEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALGVAPNDLPGEEIEPLVLG 1517
Qy   1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072
Db   1518 SECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLA 1577
Qy   1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132
Db   1578 YLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAEVYATADTPENRAYLESL 1637
Qy   1133 GVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY 1192
Db   1638 GVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLMVLRACGRLVKLGRRDDC 1697
Qy   1193 ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGV------------- 1239
Db   1698 ADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLVAAGAISPLGSGLRVGGS 1757
Qy   1240 FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHA------------- 1286
Db   1758 LTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRAPAESSVAVRADGTYLVT 1817
Qy   1287 -----------------GAG---------------------------------------- 1289
Db   1818 GGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALEAHGARVTVAKADVADRS 1877
Qy   1290 ------------------------------------------------------------ 1289
Db   1878 QIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARFRTVMGPKVQGALHLHTLTREA 1937
Qy   1290 ------------------------------------------------------------ 1289
Db   1938 PLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAQGLPALSIDWGMFTEVGMAVA 1997
Qy   1290 --------------------------------------------------PST------- 1292
Db   1998 QENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITPRQWVEFYPATAASRRLS 2057
Qy   1293 --------------GDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALF 1338
Db   2058 RLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQVSHVLRLPEDKIEVDAPL 2117
Qy   1339 TRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL-DALATAL 1389
Db   2118 SSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWLLDDALVVRL 2169
```

RESULT 9
US-09-568-102-7
; Sequence 7, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7

```
;   LENGTH: 2439
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-568-102-7

  Query Match             57.4%;  Score 4137;  DB 4;  Length 2439;
  Best Local Similarity   43.4%;  Pred. No. 8.5e-297;
  Matches  933;  Conservative  135;  Mismatches  308;  Indels  776;  Gaps   13;

Qy          6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65
Db         26 LEQERTEPIAIVGIGCRFPGGADTPEAFWELLDSGRDAV--QPLDRRWALVGVHPSEEVP 83

Qy         66 GKTPVTRASFLSD-VACPDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124
Db         84 -----RWAGLLTEAVDGFDAAFFGTSPREARSLDPQQRLLLEVTWEGLEDAGIAPQSLDG 138

Qy        125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184
Db        139 SRTGVFLGACSSDYSHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPCLTVDTA 197

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244
Db        198 CSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHCRTFDASAN 257

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304
Db        258 GFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLREAL 317

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364
Db        318 QSARVDAGAIGYVETHGTGTSLGDPIEVEALRAVLGPARADGSRCVLGAVKTNLGHLEGA 377

Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424
Db        378 AGVAGLIKAALALHHELIPRNLHFHTLNPRIRIEGTALALATEPVPWPRAGRPRFAGVSA 437

Qy        425 FGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQC 484
Db        438 FGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSAAALDAQAARLSAHIAAYPEQG 497

Qy        485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544
Db        498 LGDVAFSLVSTRSPMEHRLAVAATSREALRSALEVAAQGQTPAGAARGRAASSPGKLAFL 557

Qy        545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604
Db        558 FAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCEVMWAEPGSSRSSLLDQT 617

Qy        605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664
Db        618 AFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAGVFSLEDAVRLVVARGRL 677

Qy        665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724
Db        678 MQALPAGGAMVSIAAPEADVAAAVAPHAALVSIAAVNGPEQVVIAGAEKFVQQIAAAFAA 737

Qy        725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784
Db        738 RGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSIALVSNLSGKPCTDEVSAPGY 797

Qy        785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844
Db        798 WVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACLPDARPVLLPASRAGRDE 857

Qy        845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI----DTKADDAAR-- 898
Db        858 AASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERYWIEAPVDREADGTGRAR 917

Qy        899 ------------------------------------------------------------ 898
Db        918 AGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEVVFPGAGYLEMALSSGAE 977

Qy        899 -------------------------------------------------GDRRAP----- 904
Db        978 ILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQVASREPGERRAPFRIHA 1037

Qy        905 ------------------------------------------------------------ 904
Db       1038 RGVLRRIGRVETPARSNLAALRARLHAAVPAAAIYGALAEMGLQYGPALRGLAELWRGEG 1097

Qy        905 ------------------------------------------------------------ 904
Db       1098 EALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSP 1157

Qy        905 ------------------------------------------------------------ 904
Db       1158 GELWCHARVVSDGQQASSRWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLE 1217

Qy        905 ------------------------------------GAGH-------------------- 908
Db       1218 LDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVVHAAGDDTSTAGMRALLA 1277

Qy        909 ------------------------------------------------------------ 908
Db       1278 NAFDGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDADALESALMRGCDSVLSL 1337

Qy        909 ------------------------------------------------------------ 908
Db       1338 VQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDP 1397

Qy        909 ----------------DEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEI 952
Db       1398 AEPEGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPDAQRREKVEPAGDRPFRLEI 1457

Qy        953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLG 1012
Db       1458 DEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALGVAPNDLPGEEIEPLVLG 1517

Qy       1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072
Db       1518 SECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLA 1577

Qy       1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132
Db       1578 YLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAEVYATADTPENRAYLESL 1637

Qy       1133 GVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY 1192
Db       1638 GVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLMVLRACGRLVKLGRRDDC 1697

Qy       1193 ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGV------------- 1239
Db       1698 ADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLVAAGAISPLGSGLRVGGS 1757

Qy       1240 FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHA------------- 1286
Db       1758 LTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRAPAESSVAVRADGTYLVT 1817

Qy       1287 ----------------GAG----------------------------------------- 1289
Db       1818 GGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALEAHGARVTVAKADVADRS 1877

Qy       1290 ------------------------------------------------------------ 1289
Db       1878 QIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARFRTVMGPKVQGALHLHTLTREA 1937

Qy       1290 ------------------------------------------------------------ 1289
Db       1938 PLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAQGLPALSIDWGMFTEVGMAVA 1997

Qy       1290 -----------------------------------------------------PST------- 1292
Db       1998 QENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITPRQWVEFYPATAASRRLS 2057

Qy       1293 --------------GDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALF 1338
```

```
Db       2058 RLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQVSHVLRLPEDKIEVDAPL 2117

Qy       1339 TRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL-DALATAL 1389

Db       2118 SSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWLLDDALVVRL 2169

RESULT 10
US-09-567-969-7
; Sequence 7, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 2439
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-567-969-7

  Query Match             57.4%;  Score 4137;  DB 4;  Length 2439;
  Best Local Similarity   43.4%;  Pred. No. 8.5e-297;
  Matches  933;  Conservative  135;  Mismatches  308;  Indels  776;  Gaps   13;

Qy          6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65

Db         26 LEQERTEPIAIVGIGCRFPGGADTPEAFWELLDSGRDAV--QPLDRRWALVGVHPSEEVP 83

Qy         66 GKTPVTRASFLSD-VACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124

Db         84 -----RWAGLLTEAVDGFDAAFFGTSPREARSLDPQQRLLLEVTWEGLEDAGIAPQSLDG 138

Qy        125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184

Db        139 SRTGVFLGACSSDYSHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPCLTVDTA 197

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244

Db        198 CSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHCRTFDASAN 257

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304

Db        258 GFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLREAL 317

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364

Db        318 QSARVDAGAIGYVETHGTGTSLGDPIEVEALRAVLGPARADGSRCVLGAVKTNLGHLEGA 377

Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424

Db        378 AGVAGLIKAALALHHELIPRNLHFHTLNPRIRIEGTALALATEPVPWPRAGRPRFAGVSA 437

Qy        425 FGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQC 484

Db        438 FGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSAAALDAQAARLSAHIAAYPEQG 497

Qy        485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544

Db        498 LGDVAFSLVSTRSPMEHRLAVAATSREALRSALEVAAQGQTPAGAARGRAASSPGKLAFL 557

Qy        545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604

Db        558 FAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCEVMWAEPGSSRSSLLDQT 617

Qy        605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664

Db        618 AFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAGVFSLEDAVRLVVARGRL 677

Qy        665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724

Db        678 MQALPAGGAMVSIAAPEADVAAAVAPHAALVSIAAVNGPEQVVIAGAEKFVQQIAAAFAA 737

Qy        725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784

Db        738 RGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSIALVSNLSGKPCTDEVSAPGY 797

Qy        785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844

Db        798 WVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACLPDARPVLLPASRAGRDE 857

Qy        845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI----DTKADDAAR-- 898

Db        858 AASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERYWIEAPVDREADGTGRAR 917

Qy        899 ------------------------------------------------------------ 898

Db        918 AGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEVVFPGAGYLEMALSSGAE 977

Qy        899 -----------------------------------------------GDRRAP----- 904

Db        978 ILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQVASREPGERRAPFRIHA 1037

Qy        905 ------------------------------------------------------------ 904

Db       1038 RGVLRRIGRVETPARSNLAALRARLHAAVPAAAIYGALAEMGLQYGPALRGLAELWRGEG 1097

Qy        905 ------------------------------------------------------------ 904

Db       1098 EALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSP 1157

Qy        905 ------------------------------------------------------------ 904

Db       1158 GELWCHARVVSDGQQASSRWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLE 1217

Qy        905 ---------------------------------------GAGH------------------ 908

Db       1218 LDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVVHAAGDDTSTAGMRALLA 1277

Qy        909 ------------------------------------------------------------ 908

Db       1278 NAFDGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDADALESALMRGCDSVLSL 1337

Qy        909 ------------------------------------------------------------ 908

Db       1338 VQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDP 1397

Qy        909 ---------------DEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEI 952

Db       1398 AEPEGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPDAQRREKVEPAGDRPFRLEI 1457

Qy        953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012

Db       1458 DEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALGVAPNDLPGEEIEPLVLG 1517

Qy       1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072

Db       1518 SECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLA 1577

Qy       1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132

Db       1578 YLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAEVYATADTPENRAYLESL 1637

Qy       1133 GVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY 1192
```

```
Db       1638 GVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLMVLRACGRLVKLGRRDDC 1697

Qy       1193 ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGV------------- 1239

Db       1698 ADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLVAAGAISPLGSGLRVGGS 1757

Qy       1240 FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHA------------- 1286

Db       1758 LTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRAPAESSVAVRADGTYLVT 1817

Qy       1287 ----------------GAG----------------------------------------- 1289

Db       1818 GGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALEAHGARVTVAKADVADRS 1877

Qy       1290 ------------------------------------------------------------ 1289

Db       1878 QIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARFRTVMGPKVQGALHLHTLTREA 1937

Qy       1290 ------------------------------------------------------------ 1289

Db       1938 PLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAQGLPALSIDWGMFTEVGMAVA 1997

Qy       1290 ---------------------------------------------------PST------- 1292

Db       1998 QENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITPRQWVEFYPATAASRRLS 2057

Qy       1293 --------------GDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALF 1338

Db       2058 RLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQVSHVLRLPEDKIEVDAPL 2117

Qy       1339 TRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL-DALATAL 1389

Db       2118 SSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWLLDDALVVRL 2169

RESULT 11
US-09-568-480-7
; Sequence 7, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 2439
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-568-480-7

  Query Match             57.4%;  Score 4137;  DB 4;  Length 2439;
  Best Local Similarity   43.4%;  Pred. No. 8.5e-297;
  Matches  933;  Conservative  135;  Mismatches  308;  Indels  776;  Gaps   13;

Qy          6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65

Db         26 LEQERTEPIAIVGIGCRFPGGADTPEAFWELLDSGRDAV--QPLDRRWALVGVHPSEEVP 83

Qy         66 GKTPVTRASFLSD-VACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124

Db         84 -----RWAGLLTEAVDGFDAAFFGTSPREARSLDPQQRLLLEVTWEGLEDAGIAPQSLDG 138

Qy        125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184

Db        139 SRTGVFLGACSSDYSHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPCLTVDTA 197

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244

Db        198 CSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHCRTFDASAN 257

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304

Db        258 GFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLREAL 317

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364

Db        318 QSARVDAGAIGYVETHGTGTSLGDPIEVEALRAVLGPARADGSRCVLGAVKTNLGHLEGA 377

Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424

Db        378 AGVAGLIKAALALHHELIPRNLHFHTLNPRIRIEGTALALATEPVPWPRAGRPRFAGVSA 437

Qy        425 FGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQC 484

Db        438 FGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSAAALDAQAARLSAHIAAYPEQG 497

Qy        485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544

Db        498 LGDVAFSLVSTRSPMEHRLAVAATSREALRSALEVAAQGQTPAGAARGRAASSPGKLAFL 557

Qy        545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604

Db        558 FAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCEVMWAEPGSSRSSLLDQT 617

Qy        605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664

Db        618 AFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAGVFSLEDAVRLVVARGRL 677

Qy        665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724

Db        678 MQALPAGGAMVSIAAPEADVAAAVAPHAALVSIAAVNGPEQVVIAGAEKFVQQIAAAFAA 737

Qy        725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784

Db        738 RGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSIALVSNLSGKPCTDEVSAPGY 797

Qy        785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844

Db        798 WVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACLPDARPVLLPASRAGRDE 857

Qy        845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI----DTKADDAAR-- 898

Db        858 AASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERYWIEAPVDREADGTGRAR 917

Qy        899 ------------------------------------------------------------ 898

Db        918 AGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEVVPPGAGYLEMALSSGAE 977

Qy        899 ------------------------------------------------GDRRAP----- 904

Db        978 ILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQVASREPGERRAPFRIHA 1037

Qy        905 ------------------------------------------------------------ 904

Db       1038 RGVLRRIGRVETPARSNLAALRARLHAAVPAAIYGALAEMGLQYGPALRGLAELWRGEG 1097

Qy        905 ------------------------------------------------------------ 904

Db       1098 EALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSP 1157

Qy        905 ------------------------------------------------------------ 904

Db       1158 GELWCHARVVSDGQQASSRWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLE 1217

Qy        905 -----------------------------------GAGH-------------------- 908
```

```
Db       1218 LDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVVHAAGDDTSTAGMRALLA 1277

Qy        909 ------------------------------------------------------------ 908

Db       1278 NAFDGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDADALESALMRGCDSVLSL 1337

Qy        909 ------------------------------------------------------------ 908

Db       1338 VQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDP 1397

Qy        909 ----------------DEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEI 952

Db       1398 AEPEGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPDAQRREKVEPAGDRPFRLEI 1457

Qy        953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012

Db       1458 DEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALGVAPNDLPGEEIEPLVLG 1517

Qy       1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072

Db       1518 SECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLA 1577

Qy       1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132

Db       1578 YLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAEVYATADTPENRAYLESL 1637

Qy       1133 GVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY 1192

Db       1638 GVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLMVLRACGRLVKLGRRDDC 1697

Qy       1193 ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGV------------- 1239

Db       1698 ADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLVAAGAISPLGSGLRVGGS 1757

Qy       1240 FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHA------------- 1286

Db       1758 LTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRAPAESSVAVRADGTYLVT 1817

Qy       1287 -----------------GAG---------------------------------------- 1289

Db       1818 GGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALEAHGARVTVAKADVADRS 1877

Qy       1290 ------------------------------------------------------------ 1289

Db       1878 QIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARFRTVMGPKVQGALHLHTLTREA 1937

Qy       1290 ------------------------------------------------------------ 1289

Db       1938 PLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAQGLPALSIDWGMFTEVGMAVA 1997

Qy       1290 ----------------------------------------------------PST------- 1292

Db       1998 QENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITPRQWVEFYPATAASRRLS 2057

Qy       1293 --------------GDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALF 1338

Db       2058 RLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQVSHVLRLPEDKIEVDAPL 2117

Qy       1339 TRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL-DALATAL 1389

Db       2118 SSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWLLDDALVVRL 2169
```

RESULT 12
US-09-568-486-7
; Sequence 7, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 2439
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-568-486-7

```
  Query Match             57.4%;  Score 4137;  DB 4;  Length 2439;
  Best Local Similarity   43.4%;  Pred. No. 8.5e-297;
  Matches  933;  Conservative  135;  Mismatches  308;  Indels  776;  Gaps   13;

Qy          6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65

Db         26 LEQERTEPIAIVGIGCRFPGGADTPEAFWELLDSGRDAV--QPLDRRWALVGVHPSEEVP 83

Qy         66 GKTPVTRASFLSD-VACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124

Db         84 -----RWAGLLTEAVDGFDAAFFGTSPREARSLDPQQRLLLEVTWEGLEDAGIAPQSLDG 138

Qy        125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184

Db        139 SRTGVFLGACSSDYSHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPCLTVDTA 197

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244

Db        198 CSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHCRTFDASAN 257

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304

Db        258 GFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLREAL 317

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364

Db        318 QSARVDAGAIGYVETHGTGTSLGDPIEVEALRAVLGPARADGSRCVLGAVKTNLGHLEGA 377

Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424

Db        378 AGVAGLIKAALALHHELIPRNLHFHTLNPRIRIEGTALALATEPVPWPRAGRPRFAGVSA 437

Qy        425 FGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQC 484

Db        438 FGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSAAALDAQAARLSAHIAAYPEQG 497

Qy        485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544

Db        498 LGDVAFSLVSTRSPMEHRLAVAATSREALRSALEVAAQGQTPAGAARGRAASSPGKLAFL 557

Qy        545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604

Db        558 FAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCEVMWAEPGSSRSSLLDQT 617

Qy        605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664

Db        618 AFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAGVFSLEDAVRLVVARGRL 677

Qy        665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724

Db        678 MQALPAGGAMVSIAAPEADVAAAVAPHAALVSIAAVNGPEQVVIAGAEKFVQQIAAAFAA 737

Qy        725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784

Db        738 RGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSIALVSNLSGKPCTDEVSAPGY 797

Qy        785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844
```

```
Db      798 WVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACLPDARPVLLPASRAGRDE 857

Qy      845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI----DTKADDAAR-- 898

Db      858 AASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERYWIEAPVDREADGTGRAR 917

Qy      899 ------------------------------------------------------------ 898

Db      918 AGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEVVFPGAGYLEMALSSGAE 977

Qy      899 ------------------------------------------------GDRRAP----- 904

Db      978 ILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQVASREPGERRAPFRIHA 1037

Qy      905 ------------------------------------------------------------ 904

Db     1038 RGVLRRIGRVETPARSNLAALRARLHAAVPAAAIYGALAEMGLQYGPALRGLAELWRGEG 1097

Qy      905 ------------------------------------------------------------ 904

Db     1098 EALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSP 1157

Qy      905 ------------------------------------------------------------ 904

Db     1158 GELWCHARVVSDGQQASSRWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLE 1217

Qy      905 ------------------------------------GAGH-------------------- 908

Db     1218 LDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVVHAAGDDTSTAGMRALLA 1277

Qy      909 ------------------------------------------------------------ 908

Db     1278 NAFDGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDADALESALMRGCDSVLSL 1337

Qy      909 ------------------------------------------------------------ 908

Db     1338 VQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDP 1397

Qy      909 ----------------DEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEI 952

Db     1398 AEPEGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPDAQRREKVEPAGDRPFRLEI 1457

Qy      953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012

Db     1458 DEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALGVAPNDLPGEEIEPLVLG 1517

Qy     1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072

Db     1518 SECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLA 1577

Qy     1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132

Db     1578 YLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAEVYATADTPENRAYLESL 1637

Qy     1133 GVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY 1192

Db     1638 GVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLMVLRACGRLVKLGRRDDC 1697

Qy     1193 ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGV------------- 1239

Db     1698 ADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLVAAGAISPLGSGLRVGGS 1757

Qy     1240 FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHA------------- 1286

Db     1758 LTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRAPAESSVAVRADGTYLVT 1817

Qy     1287 -----------------GAG--------------------------------------- 1289

Db     1818 GGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALEAHGARVTVAKADVADRS 1877

Qy     1290 ------------------------------------------------------------ 1289

Db     1878 QIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARFRTVMGPKVQGALHLHTLTREA 1937

Qy     1290 ------------------------------------------------------------ 1289

Db     1938 PLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAQGLPALSIDWGMFTEVGMAVA 1997

Qy     1290 -------------------------------------------------PST------- 1292

Db     1998 QENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITPRQWVEFYPATAASRRLS 2057

Qy     1293 --------------GDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALF 1338

Db     2058 RLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQVSHVLRLPEDKIEVDAPL 2117

Qy     1339 TRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL-DALATAL 1389

Db     2118 SSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWLLDDALVVRL 2169
```

RESULT 13
US-09-568-472-7
; Sequence 7, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-7

```
  Query Match             57.4%;  Score 4137;  DB 4;  Length 2439;
  Best Local Similarity   43.4%;  Pred. No. 8.5e-297;
  Matches  933;  Conservative  135;  Mismatches  308;  Indels  776;  Gaps  13;

Qy        6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65

Db       26 LEQERTEPIAIVGIGCRFPGGADTPEAFWELLDSGRDAV--QPLDRRWALVGVHPSEEVP 83

Qy       66 GKTPVTRASFLSD-VACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124

Db       84 -----RWAGLLTEAVDGFDAAFFGTSPREARSLDPQQRLLLEVTWEGLEDAGIAPQSLDG 138

Qy      125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184

Db      139 SRTGVFLGACSSDYSHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPCLTVDTA 197

Qy      185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244

Db      198 CSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHCRTFDASAN 257

Qy      245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304

Db      258 GFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLREAL 317

Qy      305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364

Db      318 QSARVDAGAIGYVETHGTGTSLGDPIEVEALRAVLGPARADGSRCVLGAVKTNLGHLEGA 377

Qy      365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424
```

```
Db        378 AGVAGLIKAALALHHELIPRNLHFHTLNPRIRIEGTALALATEPVPWPRAGRPRFAGVSA 437

Qy        425 FGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQC 484

Db        438 FGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSAAALDAQAARLSAHIAAYPEQG 497

Qy        485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544

Db        498 LGDVAFSLVSTRSPMEHRLAVAATSREALRSALEVAAQGQTPAGAARGRAASSPGKLAFL 557

Qy        545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604

Db        558 FAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCEVMWAEPGSSRSSLLDQT 617

Qy        605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664

Db        618 AFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAGVFSLEDAVRLVVARGRL 677

Qy        665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724

Db        678 MQALPAGGAMVSIAAPEADVAAAVAPHAALVSIAAVNGPEQVVIAGAEKFVQQIAAAFAA 737

Qy        725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784

Db        738 RGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSIALVSNLSGKPCTDEVSAPGY 797

Qy        785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844

Db        798 WVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACLPDARPVLLPASRAGRDE 857

Qy        845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI----DTKADDAAR-- 898

Db        858 AASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERYWIEAPVDREADGTGRAR 917

Qy        899 ------------------------------------------------------------ 898

Db        918 AGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEVVFPGAGYLEMALSSGAE 977

Qy        899 -------------------------------------------------GDRRAP----- 904

Db        978 ILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQVASREPGERRAPFRIHA 1037

Qy        905 ------------------------------------------------------------ 904

Db       1038 RGVLRRIGRVETPARSNLAALRARLHAAVPAAAIYGALAEMGLQYGPALRGLAELWRGEG 1097

Qy        905 ------------------------------------------------------------ 904

Db       1098 EALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSP 1157

Qy        905 ------------------------------------------------------------ 904

Db       1158 GELWCHARVVSDGQQASSRWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLE 1217

Qy        905 ------------------------------------GAGH-------------------- 908

Db       1218 LDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVVHAAGDDTSTAGMRALLA 1277

Qy        909 ------------------------------------------------------------ 908

Db       1278 NAFDGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDADALESALMRGCDSVLSL 1337

Qy        909 ------------------------------------------------------------ 908

Db       1338 VQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDP 1397

Qy        909 ----------------DEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEI 952

Db       1398 AEPEGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPDAQRREKVEPAGDRPFRLEI 1457

Qy        953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012

Db       1458 DEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALGVAPNDLPGEEIEPLVLG 1517

Qy       1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072

Db       1518 SECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLA 1577

Qy       1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132

Db       1578 YLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAEVYATADTPENRAYLESL 1637

Qy       1133 GVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY 1192

Db       1638 GVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLMVLRACGRLVKLGRRDDC 1697

Qy       1193 ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGV------------ 1239

Db       1698 ADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLVAAGAISPLGSGLRVGGS 1757

Qy       1240 FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHA------------ 1286

Db       1758 LTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRAPAESSVAVRADGTYLVT 1817

Qy       1287 -----------------GAG---------------------------------------- 1289

Db       1818 GGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALEAHGARVTVAKADVADRS 1877

Qy       1290 ------------------------------------------------------------ 1289

Db       1878 QIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARFRTVMGPKVQGALHLHTLTREA 1937

Qy       1290 ------------------------------------------------------------ 1289

Db       1938 PLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAQGLPALSIDWGMFTEVGMAVA 1997

Qy       1290 ---------------------------------------------------PST------- 1292

Db       1998 QENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITPRQWVEFYPATAASRRLS 2057

Qy       1293 --------------GDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALF 1338

Db       2058 RLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQVSHVLRLPEDKIEVDAPL 2117

Qy       1339 TRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL-DALATAL 1389

Db       2118 SSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWLLDDALVVRL 2169
```

```
RESULT 14
US-09-567-899-7
; Sequence 7, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 2439
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-567-899-7

  Query Match            57.4%;  Score 4137;  DB 4;  Length 2439;
```

```
Best Local Similarity   43.4%;  Pred. No. 8.5e-297;
Matches  933;  Conservative  135;  Mismatches  308;  Indels  776;  Gaps   13;

Qy          6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65
Db         26 LEQERTEPIAIVGIGCRFPGGADTPEAFWELLDSGRDAV--QPLDRRWALVGVHPSEEVP 83
Qy         66 GKTPVTRASFLSD-VACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124
Db         84 -----RWAGLLTEAVDGFDAAFFGTSPREARSLDPQQRLLLEVTWEGLEDAGIAPQSLDG 138
Qy        125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184
Db        139 SRTGVFLGACSSDYSHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPCLTVDTA 197
Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244
Db        198 CSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHCRTFDASAN 257
Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304
Db        258 GFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLREAL 317
Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364
Db        318 QSARVDAGAIGYVETHGTGTSLGDPIEVEALRAVLGPARADGSRCVLGAVKTNLGHLEGA 377
Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424
Db        378 AGVAGLIKAALALHHELIPRNLHFHTLNPRIRIEGTALALATEPVPWPRAGRPRFAGVSA 437
Qy        425 FGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQC 484
Db        438 FGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSAAALDAQAARLSAHIAAYPEQG 497
Qy        485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544
Db        498 LGDVAFSLVSTRSPMEHRLAVAATSREALRSALEVAAQGQTPAGAARGRAASSPGKLAFL 557
Qy        545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604
Db        558 FAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCEVMWAEPGSSRSSLLDQT 617
Qy        605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664
Db        618 AFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAGVFSLEDAVRLVVARGRL 677
Qy        665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724
Db        678 MQALPAGGAMVSIAAPEADVAAAVAPHAALVSIAAVNGPEQVVIAGAEKFVQQIAAAFAA 737
Qy        725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784
Db        738 RGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSIALVSNLSGKPCTDEVSAPGY 797
Qy        785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844
Db        798 WVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACLPDARPVLLPASRAGRDE 857
Qy        845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI----DTKADDAAR-- 898
Db        858 AASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERYWIEAPVDREADGTGRAR 917
Qy        899 ------------------------------------------------------------ 898
Db        918 AGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEVVFPGAGYLEMALSSGAE 977
Qy        899 ------------------------------------------------GDRRAP----- 904
Db        978 ILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQVASREPGERRAPFRIHA 1037
Qy        905 ------------------------------------------------------------ 904
Db       1038 RGVLRRIGRVETPARSNLAALRARLHAAVPAAAIYGALAEMGLQYGPALRGLAELWRGEG 1097
Qy        905 ------------------------------------------------------------ 904
Db       1098 EALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSP 1157
Qy        905 ------------------------------------------------------------ 904
Db       1158 GELWCHARVVSDGQQASSRWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLE 1217
Qy        905 ------------------------------------GAGH-------------------- 908
Db       1218 LDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVVHAAGDDTSTAGMRALLA 1277
Qy        909 ------------------------------------------------------------ 908
Db       1278 NAFDGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDADALESALMRGCDSVLSL 1337
Qy        909 ------------------------------------------------------------ 908
Db       1338 VQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDP 1397
Qy        909 ----------------DEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEI 952
Db       1398 AEPEGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPDAQRREKVEPAGDRPFRLEI 1457
Qy        953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012
Db       1458 DEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALGVAPNDLPGEEIEPLVLG 1517
Qy       1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072
Db       1518 SECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLA 1577
Qy       1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132
Db       1578 YLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAEVYATADTPENRAYLESL 1637
Qy       1133 GVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY 1192
Db       1638 GVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLMVLRACGRLVKLGRRDDC 1697
Qy       1193 ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGV------------- 1239
Db       1698 ADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLVAAGAISPLGSGLRVGGS 1757
Qy       1240 FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHA------------- 1286
Db       1758 LTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRAPAESSVAVRADGTYLVT 1817
Qy       1287 -----------------GAG---------------------------------------- 1289
Db       1818 GGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALEAHGARVTVAKADVADRS 1877
Qy       1290 ------------------------------------------------------------ 1289
Db       1878 QIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARFRTVMGPKVQGALHLHTLTREA 1937
Qy       1290 ------------------------------------------------------------ 1289
Db       1938 PLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAQGLPALSIDWGMFTEVGMAVA 1997
Qy       1290 -----------------------------------------------------PST------- 1292
Db       1998 QENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITPRQWVEFYPATAASRRLS 2057
Qy       1293 --------------GDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALF 1338
Db       2058 RLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQVSHVLRLPEDKIEVDAPL 2117
Qy       1339 TRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL-DALATAL 1389
Db       2118 SSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWLLDDALVVRL 2169
```

```
RESULT 15
US-09-413-814-70
; Sequence 70, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
;  APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
;  APPLICANT: Bristol-Myers Squibb, Co.
;  APPLICANT: Beyer, Stefan
;  APPLICANT: Bloecker, Helmut
;  APPLICANT: Brandt, Petra
;  APPLICANT: Cino, Paul M
;  APPLICANT: Dougherty, Brian A
;  APPLICANT: Goldberg, Steven L
;  APPLICANT: Hofle, Gerhard
;  APPLICANT: Mueller, Joachim
;  APPLICANT: Reichenbach, Hans
;  TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
;  TITLE OF INVENTION: heteropolyketide compounds
;  FILE REFERENCE: PCT/US 99/23535
;  CURRENT APPLICATION NUMBER: US/09/413,814
;  CURRENT FILING DATE: 1999-10-07
;  EARLIER APPLICATION NUMBER: DE 198 46 493.2
;  EARLIER FILING DATE: 1998-10-09
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 70
;   LENGTH: 2259
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-413-814-70

  Query Match            53.1%;  Score 3831.5;  DB 3;  Length 2259;
  Best Local Similarity  43.2%;  Pred. No. 3e-274;
  Matches  859;  Conservative  118;  Mismatches  245;  Indels  767;  Gaps    9;

Qy        168 ISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKT 227
Db          1 MSYTLGLQGPCLTVDTVCSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRI 60

Qy        228 RALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSG 287
Db         61 QALSPDGHCRTFDASANGFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTG 120

Qy        288 LTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVAT 347
Db        121 LMAPNVLAQEALLREALQSARVDAGAIGYVETHGTGTSLGDPIEVEALRAVLGPARADGS 180

Qy        348 PLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRA 407
Db        181 RCVLGAVKTNLGHLEGAAGVAGLIKAALALHHELIPRNLHFHTLNPRIRIEGTALALATE 240

Qy        408 RTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALD 467
Db        241 PVPWPRAGRPRFAGVSAFGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSAAALD 300

Qy        468 AQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSP 527
Db        301 AQAARLSAHIAAYPEQGLGDVAFSLVSTRSPMEHRLAVAATSREALRSALEVAAQGQTPA 360

Qy        528 GAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLRE 587
Db        361 GAARGRAASSPGKLAFLFAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCE 420

Qy        588 VMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAG 647
Db        421 VMWAEPGSSRSSLLDQTAFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAG 480

Qy        648 VFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVV 707
Db        481 VFSLEDAVRLVVARGRLMQALPAGGAMVSIAAPEADVAAAVAPHAALVSIAAVNGPEQVV 540

Qy        708 IAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLV 767
Db        541 IAGAEKFVQQIAAAFAARGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSIALV 600

Qy        768 SNLSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACM 827
Db        601 SNLSGKPCTDEVSAPGYWVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACL 660

Qy        828 PDARPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERY 887
Db        661 PDARPVLLPASRAGRDEAASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERY 720

Qy        888 WI----DTKADDAAR--------------------------------------------- 898
Db        721 WIEAPVDREADGTGRARAGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEV 780

Qy        899 ------------------------------------------------------------ 898
Db        781 VFPGAGYLEMALSSGAEILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQ 840

Qy        899 ------GDRRAP------------------------------------------------ 904
Db        841 VASREPGERRAPFRIHARGVLRRIGRVETPARSNLAALRARLHAAVPAAAIYGALAEMGL 900

Qy        905 ------------------------------------------------------------ 904
Db        901 QYGPALRGLAELWRGEGEALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEAT 960

Qy        905 ------------------------------------------------------------ 904
Db        961 PWAPVEVGSVRLFQRSPGELWCHARVVSDGQQASSRWSADFELMDGTGAVVAEISRLVVE 1020

Qy        905 ----------------------------------------------------GAGH--- 908
Db       1021 RLASGVRRRDADDWFLELDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVV 1080

Qy        909 ------------------------------------------------------------ 908
Db       1081 HAAGDDTSTAGMRALLANAFDGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDA 1140

Qy        909 ------------------------------------------------------------ 908
Db       1141 DALESALMRGCDSVLSLVQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRT 1200

Qy        909 ---------------------------------DEVEEGGAVRGGDRRSARLDHPPPESG 935
Db       1201 IALEHAELRCISVDLDPAEPEGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPDAQ 1260

Qy        936 RREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALG 995
Db       1261 RREKVEPAGDRPFRLEIDEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALG 1320

Qy        996 MVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVL 1055
Db       1321 VAPNDLPGEEIEPLVLGSECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVL 1380

Qy       1056 PRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAE 1115
Db       1381 PRPLGLSATEAAAMPLAYLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAE 1440

Qy       1116 VHATAGTPEKRAYLESLGVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFN 1175
Db       1441 VYATADTPENRAYLESLGVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLM 1500

Qy       1176 LLRSHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLI 1235
Db       1501 VLRACGRLVKLGRRDDCADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLV 1560

Qy       1236 AAGV-------------FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRI 1282
Db       1561 AAGAISPLGSGLRVGGSLTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRA 1620

Qy       1283 PTHA---------------------------GAG---------------------- 1289
```

```
                  |   :                                         |||
Db       1621 PAESSVAVRADGTYLVTGGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALE 1680

Qy       1290 ------------------------------------------------------------ 1289

Db       1681 AHGARVTVAKADVADRSQIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARFRTVM 1740

Qy       1290 ------------------------------------------------------------ 1289

Db       1741 GPKVQGALHLHTLTREAPLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAQGLP 1800

Qy       1290 ------------------------------------------------------------ 1289

Db       1801 ALSIDWGMFTEVGMAVAQENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITP 1860

Qy       1290 -------PST---------------------GDRDLLDRLASAAPAARAAALEAFLRTQV 1321
                     |:|                     ||||||::||||  |:|||   |:   :| ||
Db       1861 RQWVEFYPATAASRRLSRLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQV 1920

Qy       1322 SQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNL 1381
              | ||| || |:  :|  : :||||||::||||||||:| :         | | :| : : |
Db       1921 SHVLRLPEDKIEVDAPLSSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWL 1980

Qy       1382 L-DALATAL 1389
              | |||   |
Db       1981 LDDALVVRL 1989
```

Search completed: October 2, 2003, 17:35:52
Job time : 71.7527 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          October  2, 2003, 17:27:29 ; Search time 95.763 Seconds
                                          (without alignments)
                                          2355.298 Million cell updates/sec

Title:           US-10-014-717-2
Perfect score:   7210
Sequence:        1 VADRPIERAAEDPIAIVGAS..........GVQNDFVSSGADQDWEIIAL 1421

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters:      1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_19Jun03:*
                 1:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
                 2:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
                 3:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
                 4:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
                 5:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
                 6:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
                 7:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
                 8:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
                 9:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
                 10:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
                 11:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
                 12:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
                 13:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
                 14:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
                 15:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
                 16:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
                 17:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
                 18:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
                 19:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
                 20:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
                 21:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
                 22:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
                 23:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
                 24:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7210	100.0	1421	21	AAY58573	Sorangium cellulos
2	4137	57.4	2439	21	AAY58578	Sorangium cellulos
3	3311.5	45.9	7257	21	AAY58576	Sorangium cellulos
4	2531.5	35.1	2153	22	AAB83974	Amino acid sequenc
5	2363	32.8	4038	22	ABG99871	S. cinnamonensis M
6	2355.5	32.7	5435	22	AAE10145	Streptomyces nours
7	2351	32.6	4106	22	ABG99872	S. cinnamonensis M
8	2331.5	32.3	11096	22	AAE10129	Streptomyces nours
9	2319.5	32.2	2149	24	ABP57679	Saccharopolyspora
10	2273	31.5	2152	20	AAY39298	SpnB a polyketide
11	2273	31.5	2152	22	AAB70966	S. spinosa protein
12	2260.5	31.4	7068	22	AAE10142	Streptomyces nours
13	2260.5	31.4	9477	22	AAE10144	Streptomyces nours
14	2250	31.2	6095	22	AAU10701	Amino acid sequenc
15	2245.5	31.1	3798	21	AAY58577	Sorangium cellulos
16	2194	30.4	5087	22	AAU10700	Amino acid sequenc
17	2165	30.0	6797	22	AAB31558	Pimaricin biosynth
18	2150	29.8	1841	18	AAW22605	Tylactone synthase
19	2135	29.6	4132	22	ABG99870	S. cinnamonensis M
20	2128	29.5	3816	21	AAY92708	S. antibioticus 8,
21	2127	29.5	3170	20	AAY39299	SpnC a polyketide
22	2127	29.5	3170	22	AAB70967	S. spinosa protein
23	2125	29.5	4630	21	AAY77177	S. venezuelae vep
24	2119	29.4	4933	24	ABP57681	Saccharopolyspora
25	2113	29.3	4924	22	AAB70968	S. spinosa protein
26	2113	29.3	4928	20	AAY39300	SpnD a polyketide
27	2108.5	29.2	3562	22	AAB82213	Polyketide synthas
28	2107.5	29.2	3567	14	AAR44431	eryA region polype
29	2104	29.2	2238	22	ABG99869	S. cinnamonensis M
30	2103	29.2	4630	18	AAW19629	Streptomyces venez
31	2097.5	29.1	1366	22	AAE10127	Streptomyces nours
32	2097	29.1	3167	24	ABP57680	Saccharopolyspora
33	2091.5	29.0	5588	20	AAY39301	SpnE a polyketide
34	2091.5	29.0	5588	22	AAB70969	S. spinosa protein
35	2078	28.8	4344	24	ABP57678	Saccharopolyspora
36	2065.5	28.6	5564	24	ABP57682	Saccharopolyspora
37	2063.5	28.6	3413	19	AAW52849	A. mediterranei ri
38	2052.5	28.5	3739	21	AAY77193	S. venezuelae macr
39	2052.5	28.5	3739	21	AAY77201	S. venezuelae pik
40	2052.5	28.5	12199	21	AAY77180	S. venezuelae pik
41	2051.5	28.5	4551	21	AAB18637	Amino acid sequenc
42	2051.5	28.5	4551	21	AAY67201	Narbonolide syntha
43	2051.5	28.5	4551	24	ABG71661	S. venezuelae narb
44	2051.5	28.5	4613	21	AAY77192	S. venezuelae macr
45	2051.5	28.5	4613	21	AAY77200	S. venezuelae pik

ALIGNMENTS

```
RESULT 1
AAY58573
ID   AAY58573 standard; Protein; 1421 AA.
XX
AC   AAY58573;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum type I polyketide synthase EPOS A.
XX
KW   EPOS A; type I polyketide synthase; epothilone biosynthesis;
KW   thiazole ring formation; taxol substitute; anticancer.
XX
OS   Sorangium cellulosum.
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   N-PSDB; AAZ55887.
```

```
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy  -
XX
PS   Claim 12; Page 105-109; 174pp; English.
XX
CC   This sequence represents a Sorangium cellulosum type I polyketide
CC   synthase, EPOS A, which is one of several epothilone biosynthetic
CC   enzymes encoded by a 68.75 kb contig. Epothilones A and B are
CC   16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence   1421 AA;

  Query Match             100.0%;  Score 7210;  DB 21;  Length 1421;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1421;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDP 60

Qy         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120

Qy        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180

Qy        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240

Qy        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300

Qy        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360

Qy        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRA 420

Qy        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETY 480

Qy        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 PSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGK 540

Qy        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 LAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAAL 600

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660

Qy        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720

Qy        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVS 780

Qy        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 SPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRA 840

Qy        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGD 900

Qy        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 RRAPGAGHDEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDH 960

Qy        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 LVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIV 1020

Qy       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYAL 1080

Qy       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 DRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDS 1140

Qy       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 RSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR 1200

Qy       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 PFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260

Qy       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ 1320

Qy       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQN 1380

Qy       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421
              |||||||||||||||||||||||||||||||||||||||||
Db       1381 LLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL 1421


RESULT 2
AAY58578
ID   AAY58578 standard; Protein; 2439 AA.
XX
AC   AAY58578;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum type I polyketide synthase EPOS E.
XX
KW   EPOS E; type I polyketide synthase; epothilone biosynthesis;
```

```
KW   polyketide backbone formation; taxol substitute; anticancer.
XX
OS   Sorangium cellulosum.
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   N-PSDB; AAZ55887.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy  -
XX
PS   Claim 12; Page 151-158; 174pp; English.
XX
CC   This sequence represents a Sorangium cellulosum type I polyketide
CC   synthase, EPOS E, which is one of several epothilone biosynthetic
CC   enzymes encoded by a 68.75 kb contig. Epothilones A and B are
CC   16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence   2439 AA;

  Query Match             57.4%;  Score 4137;  DB 21;  Length 2439;
  Best Local Similarity   43.4%;  Pred. No. 0;
  Matches  933;  Conservative  135;  Mismatches  308;  Indels  776;  Gaps   13;

Qy          6 IERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAP 65

Db         26 LEQERTEPIAIVGIGCRFPGGADTPEAFWELLDSGRDAV--QPLDRRWALVGVHPSEEVP 83

Qy         66 GKTPVTRASFLSD-VACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVG 124

Db         84 -----RWAGLLTEAVDGFDAAFFGTSPREARSLDPQQRLLLEVTWEGLEDAGIAPQSLDG 138

Qy        125 TETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTA 184

Db        139 SRTGVFLGACSSDYSHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPCLTVDTA 197

Qy        185 YSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEAD 244

Db        198 CSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHCRTFDASAN 257

Qy        245 GFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRAL 304

Db        258 GFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLREAL 317

Qy        305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364

Db        318 QSARVDAGAIGYVETHGTGTSLGDPIEVEALRAVLGPARADGSRCVLGAVKTNLGHLEGA 377

Qy        365 SGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSS 424

Db        378 AGVAGLIKAALALHHELIPRNLHFHTLNPRIRIEGTALALATEPVPWPRAGRPRFAGVSA 437

Qy        425 FGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQC 484

Db        438 FGLSGTNVHVVLEEAPATVLAPATPGRSAELLVLSAKSAAALDAQAARLSAHIAAYPEQG 497

Qy        485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544

Db        498 LGDVAFSLVSTRSPMEHRLAVAATSREALRSALEVAAQGQTPAGAARGRAASSPGKLAFL 557

Qy        545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604

Db        558 FAGQGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQPLCEVMWAEPGSSRSSLLDQT 617

Qy        605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664

Db        618 AFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAGVFSLEDAVRLVVARGRL 677

Qy        665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724

Db        678 MQALPAGGAMVSIAAPEADVAAAVAPHAALVSIAAVNGPEQVVIAGAEKFVQQIAAAFAA 737

Qy        725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784

Db        738 RGARTKPLHVSHAFHSPLMDPMLEAFRRVTESVTYRRPSIALVSNLSGKPCTDEVSAPGY 797

Qy        785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844

Db        798 WVRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACLPDARPVLLPASRAGRDE 857

Qy        845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWI----DTKADDAAR-- 898

Db        858 AASALEALGGFWVVGGSVTWSGVFPSGGRRVPLPTYPWQRERYWIEAPVDREADGTGRAR 917

Qy        899 ------------------------------------------------------------ 898

Db        918 AGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEVVPPGAGYLEMALSSGAE 977

Qy        899 ---------------------------------------------GDRRAP------ 904

Db        978 ILGDGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRFQVASREPGERRAPFRIHA 1037

Qy        905 ------------------------------------------------------------ 904

Db       1038 RGVLRRIGRVETPARSNLAALRARLHAAVPAAAIYGALAEMGLQYGPALRGLAELWRGEG 1097

Qy        905 ------------------------------------------------------------ 904

Db       1098 EALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSP 1157

Qy        905 ------------------------------------------------------------ 904

Db       1158 GELWCHARVVSDGQQASSRWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLE 1217

Qy        905 ----------------------------------GAGH-------------------- 908

Db       1218 LDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVVHAAGDDTSTAGMRALLA 1277

Qy        909 ------------------------------------------------------------ 908

Db       1278 NAFDGQAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDADALESALMRGCDSVLSL 1337
```

```
Qy        909 ------------------------------------------------------------ 908

Db       1338 VQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDP 1397

Qy        909 ----------------DEVEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEI 952

Db       1398 AEPEGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPDAQRREKVEPAGDRPFRLEI 1457

Qy        953 DEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLG 1012

Db       1458 DEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALGVAPNDLPGEEIEPLVLG 1517

Qy       1013 GECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVA 1072

Db       1518 SECAGRIVAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLA 1577

Qy       1073 YLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESL 1132

Db       1578 YLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAEVYATADTPENRAYLESL 1637

Qy       1133 GVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCY 1192

Db       1638 GVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLMVLRACGRLVKLGRRDDC 1697

Qy       1193 ADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGV------------- 1239

Db       1698 ADTQPGLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLVAAGAISPLGSGLRVGGS 1757

Qy       1240 FTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHA------------- 1286

Db       1758 LTPPPVETFPISRAAEAFRRMAQGQHLGKLVLTLDDPEVRIRAPAESSVAVRADGTYLVT 1817

Qy       1287 -----------------GAG---------------------------------------- 1289

Db       1818 GGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVAALEAHGARVTVAKADVADRS 1877

Qy       1290 ------------------------------------------------------------ 1289

Db       1878 QIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTPARFRTVMGPKVQGALHLHTLTREA 1937

Qy       1290 ------------------------------------------------------------ 1289

Db       1938 PLSFFVLYASAAGLFGSPGQGNYAAANAFLDALSHHRRAQGLPALSIDWGMFTEVGMAVA 1997

Qy       1290 ----------------------------------------------------PST------- 1292

Db       1998 QENRGARQISRGMRGITPDEGLSALARLLEGDRVQTGVIPITPRQWVEFYPATAASRRLS 2057

Qy       1293 --------------GDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALF 1338

Db       2058 RLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQVSHVLRLPEDKIEVDAPL 2117

Qy       1339 TRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLL-DALATAL 1389

Db       2118 SSMGMDSLMSLELRNRIEAALGVAAPAALGWTYPTVAAITRWLLDDALVVRL 2169
```

RESULT 3
AAY58576
ID   AAY58576 standard; Protein; 7257 AA.
XX
AC   AAY58576;
XX
DT   10-APR-2000 (first entry)
XX
DE   Sorangium cellulosum type I polyketide synthase EPOS C.
XX
KW   EPOS C; type I polyketide synthase; epothilone biosynthesis;
KW   polyketide backbone formation; taxol substitute; anticancer.
XX
OS   Sorangium cellulosum.
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS) NOVARTIS AG.
PA   (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   N-PSDB; AAZ55887.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 12; Page 118-139; 174pp; English.
XX
CC   This sequence represents a Sorangium cellulosum type I polyketide
CC   synthase, EPOS C, which is one of several epothilone biosynthetic
CC   enzymes encoded by a 68.75 kb contig. Epothilones A and B are
CC   16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence   7257 AA;

```
  Query Match             45.9%;  Score 3311.5;  DB 21;  Length 7257;
  Best Local Similarity   41.1%;  Pred. No. 2.3e-254;
  Matches  784;  Conservative  162;  Mismatches  432;  Indels  531;  Gaps   25;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKT 68

Db       3021 ASDEPIAIVGAACRFPGGVEDLESYWQLLAEGVVVSAEVPADRWDAADWYDPDPEIPGRT 3080

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETG 128

Db       3081 YVTKGAFLRDLQRLDATFFRISPREAMSLDPQQRLLLEVSWEALESAGIAPDTLRDSPTG 3140

Qy        129 VFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSS 188

Db       3141 VFVGAGPNEYYTQRLRGFTDGAAGLYGGTGNMLSVTAGRLSFFLGLHGPTLAMDTACSSS 3200

Qy        189 LVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR 248

Db       3201 LVALHLACQSLRLGECDQALVGGVNVLLAPETFVLLSRMRALSPDGRCKTFSADADGYAR 3260

Qy        249 GEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAG 308
```

```
Db     3261 GEGCAVVVLKRLRDAQRAGDSILALIRGSAVNHDGPSSGLTVPNGPAQQALLRQALSQAG 3320

Qy      309 CAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368

Db     3321 VSPVDVDFVECHGTGTALGDPIEVQALSEVYGPGRSGDRPLVLGAAKANVAHLEAASGLA 3380

Qy      369 GLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMS 428

Db     3381 SLLKAVLALRHEQIPAQPELGELNPHLPWNTLPVAVPRKAVPWGRGARPRRAGVSAFGLS 3440

Qy      429 GTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDV 488

Db     3441 GTNVHVVLEEAPEVEPAPAAPARPVELVVLSAKSAAALDAAAARLSAHLSAHPELSLGDV 3500

Qy      489 AFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQ 548

Db     3501 AFSLATTRSPMEHRLAIATTSREALRGALDAAAQQKTPQGAVRGKAVSSRGKLAFLFTGQ 3560

Qy      549 GAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQ 608

Db     3561 GAQMPGMGRGLYETWPAFREAFDRCVALFDREIDQPLREVMWAAPGLAQAARLDQTAYAQ 3620

Qy      609 PALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQAL 668

Db     3621 PALFALEYALAALWRSWGVEPHVLLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMQAL 3680

Qy      669 PAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGAR 728

Db     3681 PAGGAMVAIAASEAEVAASVAPHAATVSIAAVNGPDAVVIAGAEVQVLALGATFAARGIR 3740

Qy      729 TKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRH 788

Db     3741 TKRLAVSHAFHSPLMDPMLEDFQRVAATIAYRAPDRPVVSNVTGHVAGPEIATPEYWVRH 3800

Qy      789 AREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATV 848

Db     3801 VRSAVRFGDGAKALHAAGAATFVEVGPKPVLLGLLPACLGEADAVLVPSLRADRSECEVV 3860

Qy      849 LEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDA------------ 896

Db     3861 LAALGAWYAWGGALDWKGVFPDGARRVALPMYPWQRERHWMDLTPRSAAPAGIAGRWPLA 3920

Qy      897 ------------------------------------------------------------ 896

Db     3921 GVGLCMPGAVLHHVLSIGPRHQPFLGDHLVFGKVVVPGAFHVAVILSIAAERWPERAIEL 3980

Qy      897 --------------------------ARGD---------------RR------------- 902

Db     3981 TGVEFLKAIAMEPDQEVELHAVLTPEAAGDGYLFELATLAAPETERRWTTHARGRVQPTD 4040

Qy      903 -APGA------------------------------------------------------- 906

Db     4041 GAPGALPRLEVLEDRAIQPLDFAGFLDRLSAVRIGWGPLWRWLQDGRVGDEASLATLVPT 4100

Qy      907 ---GHD----------------------EVEEGGA------------------VR-GGD 921

Db     4101 YPNAHDVAPLHPILLDNGFAVSLLSTRSEPEDDGTPPLPFAVERVRWWRAPVGRVRCGGV 4160

Qy      922 RRSA---------------------------------------------RLDHPPP--- 932

Db     4161 PRSQAFGVSSFVLVDETGEVVAEVEGFVCRRAPREVFLRQESGASTAALYRLDWPEAPLP 4220

Qy      933 --------------------------------------------------------ESGR 936

Db     4221 DAPAERIEESWVVVAAPGSEMAAALATRLNRCVLAEPKGLEAALAGVSPAGVICLWEAGA 4280

Qy      937 REKVEAAGDR-------------------------------------------------- 946

Db     4281 HEEAPAAAQRVATEGLSVVQALRDRAVRLWWVTMGAVAVEAGERVQVATAPVWGLGRTVM 4340

Qy      947 ------------------------------------------------------------ 946

Db     4341 QERPELSCTLVDLEPEADAARSADVLLRELGRADDETQVAFRSGKRRVARLVKATTPEGL 4400

Qy      947 ------PFRLEIDEPGVLDHLVLRVTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDD 1000

Db     4401 LVPDAESYRLEAGQKGTLDQLRLAPAQRRAPGPGEVEIKVTASGLNFRTVLAVLGM---- 4456

Qy     1001 LPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQA 1060

Db     4457 YPGDAGP---MGGDCAGVATAVGQGVRHVAVGDAVMTL--GTLHRFVTVDARLVVRQPAG 4511

Qy     1061 LSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATA 1120

Db     4512 LTPAQAATVPVAFLTAWLALHDLGNLRRGERVLIHAAAGGVGMAAVQIARWIGAEVFATA 4571

Qy     1121 GTPEKRAYLESLGV--RYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLR 1178

Db     4572 -SPSKWAAVQAMGVPRTHIASSRTLEFAETFRQVTGGRGVDVVLNALAGEFVDASLSLLS 4630

Qy     1179 SHGRFVELGKRDCYADNQ--LGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIA 1236

Db     4631 TGGRFLEMGKTD-IRDRAAVAAAHPGVRYRVFDILEL------APDRTREILERVVEGFA 4683

Qy     1237 AGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPT-----HAGAGP- 1290

Db     4684 AGHLRALPVHAFAITKAEAAFRFMAQARHQGKVVL-LPAPSAAPLAPTGTVLLTGGLGAL 4742

Qy     1291 -----------------STGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVG 1333

Db     4743 GLHVARWLAQQGVPHMVLTGRRG-LDTPGAAKAVAEIEALGARVTIAASDVADRNALEAV 4801

Qy     1334 AEAL---------------------------FTRLGMDSLMAVELRNRIEASLKLKLST 1365

Db     4802 LQAIPAEWPLQGVIHAAGALDDGVLDEQTTDRFSRVLAPKVTGAWNLHELTAGNDLAFFV 4861

Qy     1366 TFLSTS--------PNIALLAQNLLDALATALSLERVAAENLRAGVQND 1406

Db     4862 LFSSMSGLLGSAGQSNYA-AANTFLDALAAHRRAEGLAAQSLAWGPWSD 4909
```

RESULT 4
AAB83974
ID   AAB83974 standard; Protein; 2153 AA.
XX
AC   AAB83974;
XX
DT   06-AUG-2001 (first entry)
XX
DE   Amino acid sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase.
XX
OS   Unidentified.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 1
FT   /note= "Met encoded by GTG"
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET) AVENTIS PHARMA SA.
XX
PI   Jeannin P, Pernodet J, Guerineau M, Simonet P, Courtois S;
PI   Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.

```
DR   N-PSDB; AAF90037.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics -
XX
PS   Claim 78; Page 336-344; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAB83971-76
CC   represent type I polyketide synthases encoded by cosmid a26g1.
XX
SQ   Sequence   2153 AA;

  Query Match             35.1%;  Score 2531.5;  DB 22;  Length 2153;
  Best Local Similarity   31.4%;  Pred. No. 1.1e-192;
  Matches  665;  Conservative  187;  Mismatches  507;  Indels  761;  Gaps   25;

Qy          7 ERAAEDPIAIVGASCRLPGGVIDLSG---FWTLLEGSRDTVGRVPAERWDAAAWFDPDPD 63
Db         32 ERDAREPIAIVGLGCRFPGG----DGPDEFWQMLRSGVDAIREVPPGRWDEESVRRILKS 87

Qy         64 APGKTPV-TRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSAL 122
Db         88 LNPATPVKIQAGFLDSIDGFDNDFFGISPREAVSIDPQQRLLLEVAWEALEDAGQTMEGL 147

Qy        123 VGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVD 182
Db        148 SGSRTGVFVGIHSQSSDYFWMQTADGARIDPYTATGTAHSVIAGRLSYLLNLQGPSIALD 207

Qy        183 TAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAE 242
Db        208 TACSSSLAAVHLACQSLRSGECTLAVAGGVNLRFSPEFMYATSKMGTASPSGRCRAFDAA 267

Qy        243 ADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKR 302
Db        268 ADGIVFGEGCGVVVLKRLSDALAAGDRVWAVVRGSAVNQDGRSAGLTAPNVVSQQVVIRS 327

Qy        303 ALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPE 362
Db        328 ALANAGVAAQQIGYIEAHGTGTPLGDPIEIEALAETVGLPRPVGDVCAVGSLKSNIGHLE 387

Qy        363 YASGITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGV 422
Db        388 GAAGIAGLIKAVLALSHETIPPSLHVRQLNPNIRLEGTSLDIVKEVRPWPAGSRRRFAGV 447

Qy        423 SSFGMSGTNAHVVLEEA----------------PAATC---------------------- 444
Db        448 SAFGWSGTNAHVVLEEAAPTGRGEAASGFHSRPPAAARAAVPLAEGDTGGTPDIAGTPD 507

Qy        445 ---TPPAPERP-----------------AELLVLSARTASALDAQAARLRDHLETYPSQC 484
Db        508 TADTPDTADTPDIAGTAGTAATTGIADAMYVLPLSAHGADELRRVARAYGELLTASHAPS 567

Qy        485 LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFL 544
Db        568 LRDLCYTAAVRRTHHRCRLAVSGRTAEELAAQL----QGITIP--------SQRRKTVFV 615

Qy        545 FTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQT 604
Db        616 FSGQGSQWIGMGRSWMDREPVIREALERCEAAMRPYVDWSLKE---------ELAKLDRV 666

Qy        605 AFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRL 664
Db        667 EVIQPALFALQVAIAALWRSWGIEPDAVIGHSMGEVAAAHVAGALTLQDAARIICSRSRL 726

Qy        665 MQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAA 724
Db        727 LSRISGLGGMAMVELPLAECEAVLSTYTERLSPAVSNGPNSTVISGEVEALAEVVATLER 786

Qy        725 RGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGY 784
Db        787 RGVSCRPVKVDFAAHSPQVDPLCDELLQSLDGIQPRPATIPFYSTVTGATLETTSLDSTY 846

Qy        785 WVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDE 844
Db        847 WARNLRSPVLFWQGIRHLADSGHDVFLEISPHPILL---PAIGGNA--ALVPSLRRDQDE 901

Qy        845 PATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTK------------ 892
Db        902 RGSMLTSLGALYEAGHTVAWRTVYPS-GNCVRLPRYPWQRRRFWLDASPARHAITLGNPL 960

Qy        893 ----------------------------ADDAARGDRRAPG--------AGHDE------ 910
Db        961 LGKRVEASTQPGTFFWETELSLASVPWLADHRVQGEVVLPATAYLDMALAGTSETFGESP 1020

Qy        911 ------------------------------------------------------------ 910
Db       1021 CVLEHVTFTQMLIVPRDGSMTLQLAIAVDRPGMASFRISSRQASTWVLHASGDIRQTPAD 1080

Qy        911 ----------------------------------VEEGGAVR------------------- 918
Db       1081 ASTVPPDSAETVQARCPTVVPAAELWRQMAEHGVEYGPAFRALEQIWSCPGEAIGRLRSS 1140

Qy        919 -------------------------------GGDR------------------------- 922
Db       1141 ETRSTAPAFLDACLQIIAAAFGPAGGTWLPAGIDRMRWLHPARSVVWTHARLEGPIADLS 1200

Qy        923 ----------------------------------------------------RSAR------ 926
Db       1201 LLDGEGQLVARIEGLRLQRLDASERIDMRGWLHELRWVAQPHAAAEPPAARAARSWLIVG 1260

Qy        927 --------------------------------------LDHPP----------------- 931
Db       1261 AVDSALTAWLRATGNRVTQTSPEKLDELQPPLEEIVFLLEHEPSCDRILHLLQTLGRTPW 1320

Qy        932 -----------------------------------------------------PESGRREK 939
Db       1321 RQAPRLWLVTRGAQPVDGQILQAGIAQAPFWGLGRTVHYEHPELNCTLIDLDPAGGEEEL 1380

Qy        940 VE-----------------------------------AAGDRPFRLEIDEPGVLDHLVLR 964
Db       1381 LHELLTNNGENQIAFRGGARYVARVARHEADMQPAMFKAGDRPFRLEIDAPGVLDRLRLR 1440

Qy        965 VTERRAPGLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGE 1024
Db       1441 ATSRRPPQAGEVEIEVCAAGLNFLDVLLALGVMPDDAPGAIAGSPRLGGECSGRIVAMGK 1500

Qy       1025 GVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIA 1084
Db       1501 GVTDFRIGDEVVALAPCSFGRFVTTPAFRVALKPANIPAEQAAALPIAFLTADYALSRAA 1560

Qy       1085 RLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDR 1144
Db       1561 RLAPGERVLIHAATGGVGLAAIQIAQRAGAEIFATAGSPEKRAYLRSLGIAHVSDSRSMA 1620

Qy       1145 FVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR 1204
Db       1621 FVDDIRNWTNQEGVDVVLNSLSGDLLEASFDLLRDHGRFIEIGKRDYYAGRKLGLRPFLK 1680

Qy       1205 NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQ 1264
Db       1681 NLSYTLVDLLGMSLKRPALTRELLQEMVAKFESETWRPLETRVTTITESVEAFRTMAQAR 1740

Qy       1265 HLGKLVLTLGD------------------------------------------------- 1275
Db       1741 HIGKIVMAMRDCANAPIAPLRSAFDSEGTYLITGGLGGLGLTVARWMIGRGARRLVLLSR 1800
```

```
Qy       1276 ----PEVQ-------------------------------IRIPTHAGA----------- 1288
                  ||||                               :|   || |
Db       1801 RAPSPEVQQAIAVMDADVRTVQADVSQRDELERVISSIDRLRGVIHAAAVLDDALLLNQT 1860

Qy       1289 -----------------------------------------GPSTGD------------- 1294
                                                       |: |:
Db       1861 EAHFRNVMAAKIDGAWNLHLLTRDCPLDHFVLFSSAAGLLGAPAQGNYAAANAFLDALAY 1920

Qy       1295 ---------------------------------------------------RDLLD----- 1299
                                                                 ||:
Db       1921 YRKAQGLPALSIGWGAWSEVGLAAAQDNRGSRLALRGMENLTPQHGLAILEQLLNSSACH 1980

Qy       1300 ------------------------------------------RLASAAPAARAAALEAF 1316
                                                        || ||  |   |   ||
Db       1981 VAAMPINVRQWRQFYPKAAQSALFELLHDDAASEADAPNALRARLQSAEPQTRRTLLEEH 2040

Qy       1317 LRTQVSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIAL 1376
              |: |::::|||          :           || |||||:| |||:| :| |  |  :   | :|
Db       2041 LQQQLARVLRIDSQTIDPLRPLKELGFDSLMALEFRNRLELTLGLTLPATLIWGHPTLAG 2100

Qy       1377 LAQNLLDALATALSLERVAA 1396
              || :|   :   |   : ||
Db       2101 LAPHLASQMGLPLVEAQAAA 2120
```

RESULT 5
ABG99871
ID   ABG99871 standard; Protein; 4038 AA.
XX
AC   ABG99871;
XX
DT   16-JAN-2003 (first entry)
XX
DE   S. cinnamonensis MonAIV/polyketide synthase multi-enzyme MONS4.
XX
KW   Monensin; gene cluster; polyketide synthase; antibiotic;
KW   antihelminthic; insecticide; immunosuppressant; antifungal;
KW   antibacterial; polyether; mon BI; mon BII; mon CI; mon CII; mon H;
KW   mon RI; mon RII; mon T; mon AIX; mon AX.
XX
OS   Streptomyces cinnamonensis.
XX
PN   WO200168867-A1.
XX
PD   20-SEP-2001.
XX
PF   30-MAY-2000; 2000WO-GB02072.
XX
PR   28-MAY-1999; 99GB-0012563.
XX
PA   (BIOT-) BIOTICA TECHNOLOGY LTD.
XX
PI   Leadlay PF, Staunton J, Oliynyk M;
XX
DR   WPI; 2001-611393/70.
DR   N-PSDB; ABX04971.
XX
PT   New DNA sequence encoding polyketide synthase, useful for the
PT   production of polyketides such as antibiotic monensin -
XX
PS   Claim 6; Page 82-85; 212pp; English.
XX
CC   The invention relates to a DNA sequence which is a fully defined sequence
CC   of 103551 base pairs appearing as ABX04971, or its variant, that it is
CC   not a sequence encoding all or part amino acids 1-920 encoded by mon AI
CC   as given in the specification. The DNA is the S. cinnamonensis
CC   polyketide antibiotic monensin biosynthetic gene cluster. Also included
CC   are a recombinant cloning or expression vector comprising the gene
CC   cluster, a transformed host cell which has been transformed to
CC   contain the gene cluster (and is capable of expressing a corresponding
CC   polypeptide), a hybridization probe derived from the gene cluster
CC   (for identification and isolation of the same or analogous gene cluster,
CC   e.g. one which binds specifically to a region of the monensin gene
CC   cluster selected from mon BI, mon BII, mon CI, mon CII, mon H, Mon RI,
CC   mon RII, mon T, mon AIX and mon AX), the use of the mon RI gene or
CC   variant and a monensin promoter to control expression of a heterologous
CC   gene in Streptomyces cinnamonensis, a polypeptide encoded by a portion of
CC   the monensin gene cluster (preferably comprising mon BI, mon BII,
CC   mon AIX or mon AX or their mutants, alleles or variants), an epoxidase
CC   enzyme encoded by mon CI, a cyclase enzyme encoded by mon CII, producing
CC   S. cinnamonensis capable of enhanced levels of production of monensin
CC   comprising engineering it to overexpress the mon RI gene, S.
CC   cinnamonensis containing multiple copies of the mon RI gene and/or its
CC   variants, expressing a gene heterologous to S. cinnamonensis comprising
CC   transforming S. cinnamonensis with DNA encoding a heterologous gene and
CC   expressing the gene under control of the activator gene mon RI or
CC   actII/orf4 and 13-propyl erythromycin A. The processes and materials
CC   (enzyme systems, nucleic acids and vectors) are useful for preparing
CC   polyketides by recombinant synthesis. The polyketides are useful as
CC   insecticides, antibiotics, anthelminthics, antifungals, antibacterials or
CC   other pharmaceuticals. In particular the gene is useful for the
CC   production of monensin, an antibiotic polyether polyketide. The
CC   present sequence represents a protein encoded by the monensin gene
CC   cluster.
XX
SQ   Sequence   4038 AA;

```
  Query Match             32.8%;  Score 2363;  DB 22;  Length 4038;
  Best Local Similarity   33.6%;  Pred. No. 8.8e-179;
  Matches  647;  Conservative  193;  Mismatches  515;  Indels  572;  Gaps   47;

Qy          2 ADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDP 60
              |  ::   :||:|||  :|| ||||        ||:  :| :|  | :| ||    | |
Db       1823 AGADVDADTDDPVAIVAMTCRFPGGVASPDDLWDLLDARKDAMGAFPTDRGWDLERLFHP 1882

Qy         61 DPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPS 120
              ||| || :    :  ||    |||:||||:||||| |||  |||||  || || | |  |:
Db       1883 DPDHPGTSYTDQGGFLPDAGDFDAAFFGINPREALAMDPQQRLLLEASWEVLERAGIDPT 1942

Qy        121 ALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVA 180
               | || || ::|:    :|   :   ||  |:::  :   |  :  |:  :||::| ||| || |
Db       1943 TLKGTPTGTYVGLMYHDYAKSFP--TADAQLEGYSYLASTGSMVSGRVAYTLGLEGPAVT 2000

Qy        181 VDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFS 240
              |||| |||||::|||  |:|| |||  ||||||::|   |       |: | |: ||||||::
Db       2001 VDTACSSSLVSIHLATQALRHGECDLALAGGVTVMADPDMFAGFSRQRGLSPDGRCKAYA 2060

Qy        241 AEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVL 300
              | ||| |  ||   |::|:||| ||   | |:| |:||||:| ||||:||| ||| ||| |:
Db       2061 AAADGVGFSEGVGVLLLERLSDARRHGRRVLGVVRGSAVNQDGASNGLTAPNGPSQERVI 2120

Qy        301 KRALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
              ::|||  | ::     |  || ||||||||||| |||  || |    |     || :||||:|:||
Db       2121 RQALASGGLSSVDVDVVEGHGTGTTLGDPIEAQALLATYGQGRPEDRPLWLGSVKSNIGH 2180

Qy        361 PEYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISW--GDLRLTVTRARTPWPD-WNTP 417
               : |:|: |::|:|::::|| :|| ||     :| :  |   | :|| |      |||     |
Db       2181 TQAAAGVAGVIKMVMAMRHGVVPASLHVDVPSPHVEWDSGAVRLAV--ESVPWPQVEGRP 2238

Qy        418 RRAGVSSFGMSGTNAHVVLEEAP------AATCTPPAPERPAE----LLVLSARTASALD 467
              |||||||| ||||||::|  |        : :      | |    :         |:|||:  ||
Db       2239 RRAGVSSFGASGTNAHVIVESVPDGLEEDSVSVGGEALETETDGRLVPWVVSARSPQALR 2298

Qy        468 AQAARLRDHLETYPSQC-LGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTS 526
                || ||||    :   | || :||  ||:  |||  |      |  | |||:|  | |:
Db       2299 DQALRLRDFASDASFRAPLADVGWSLLKTRALHEHRAVVVGAERAELIAALEALATGEPH 2358

Qy        527 PGAVRSIADSSR---GKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDR 583
                |      :|     : :||:|||:| :|||:|: :  |||  |  :    |  |:
Db       2359 AALVGPACSQARVGGDDVVWLFSGQGSQLVGMGAGLYERPFPVFAAAFDEVCGL----LEG 2414

Qy        584 P-------LREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHS 636
              |        ||||::  |           ||  |  :  |   ||   :   || || || |::| |||
Db       2415 PLGVEAGGLREVVFRGPRE----RLDHTVWAQAGLFALQVGLARLWESVGVRPDVVLGHS 2470
```

```
Qy    637 IGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVS 696
Db   2471 IGEIAAAHVAGVFDLADACRVVGARARLMGGLPEGGAMCAVQATPAELAADV--DGSAVS 2528

Qy    697 IAAVNAPDQVVIAGAGQPVHAIAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAES 756
Db   2529 VAAVNTPDSTVISGPSDEVDRIAGVWRERGRKTKALSVSHAFHSALMEPMLAEFTEAIRG 2588

Qy    757 VSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPK 816
Db   2589 VKFRQPSIPLMSNVSGERAGEEITDPEYWARHVRNAVLFQPAI-AQVADSAGVFVELGPA 2647

Qy    817 STLLGLVPACM--PDARPALLASSRAG-RDEPATVLEALGGLWAVGGLVSWAGLFPSGGR 873
Db   2648 PVLTTAAQHTLDESDSQESVLVASLAGERPEESAFVEAMARLHTAGVAVDWSVLF--AGD 2705

Qy    874 RVP----LPTYPWQRERYWIDTK------------------------------------- 892
Db   2706 RVPGLVELPTYAFQRERFWLSGRSGGGDAATLGLVAAGHPLLGAAVEFADRGGCLLTGRL 2765

Qy    893 --------ADDAARGDRRAPGA--------GHDE------------------------- 910
Db   2766 SRSGVSWLADHVVAGAVLVPGAALVEWALRAGDEVGCVTVEELMLQAPLVVPEASGLRVQ 2825

Qy    911 --VEEGG----------------AVRGGDR-----------RSARLD------HPP---- 931
Db   2826 VVVEEAGEDGRRGVQIYSRPDADAVGGDDSWICHATGVLSPESARLDTELGGVWPPAGAE 2885

Qy    932 ------------------------------------------PES--------------- 934
Db   2886 PLDVDGFYAQAGEAGYGYGPAFRGLRAVWRHGQDLLAEVVLPEAAGAHDGYGIHPALLDA 2945

Qy    935 ------------------------------------------------------------ 934
Db   2946 TLHPLLAARFMDGSEDDQLYVPFGWAGVSLRAVGATTVRVRLRPVGESVDQGLSVTVTDA 3005

Qy    935 ----------------------------------------------------------GR 936
Db   3006 TGGPVLSVDSLQTRPVKPSQLAAAQQPDVRGLFTVEWTPLPQTDADGEADWVVLSDGVGR 3065

Qy    937 -REKVEAAG--------------------------------------------------- 944
Db   3066 LADVVSAAGGEAPWAVVAPVDASVGDGREGLDGRLVVERVLSLVQEFLALPELAESRLLV 3125

Qy    945 ---------------------------------DRPFRL--------------------- 950
Db   3126 VTRGAVATGVDGDGDVDASAAAVWGLVRSAQSENPGRFILLDVDGDGDDQGPDLNGRHLP 3185

Qy    951 ---------EIDEP------GVLDHLVLRVTERR-------APG---------------- 972
Db   3186 HATLRHAAEELDEPQLALREGTL--YVPRLTQARQSAELVVPPGEPAWRLRMVHDGSLDA 3243

Qy    973 -------------LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGR 1018
Db   3244 LAAVACPEALEPLAPGQVRIAVHAAGINFRDVLVALGMVP--------AYGAMGGEGAGV 3295

Qy   1019 IVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWY 1078
Db   3296 VTEVGPEVTHVSVGDRVMGVFEGAFGPVVIAEARMVTPVPQGWDMREAAGIPAAFLTAWY 3355

Qy   1079 ALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGV--RY 1136
Db   3356 GLVELAGLKAGERVLVHAATGGVGMAAVQIARHVGAEVFATA-SPGKHAVLEEMGIDAAH 3414

Qy   1137 VSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQ 1196
Db   3415 RASSRDLAFEGTFREATGGRGMDVVLNSLAGEFIDASLRLLGDGGRFLEMGKTDVRAAEE 3474

Qy   1197 LGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADA 1256
Db   3475 VAAEH--ADVSYTAYDLVGD--AGPDRISNMLDKLVELFASERLKPLPVRSWPLDKAQEA 3530

Qy   1257 FRSMAQAQHLGKLVLTLG---DPEVQIRIPTHAGAGPST---------GDRDLL--DRLA 1302
Db   3531 FRFMSQAKHTGKLVLEIPPALDPEGTVLVTGGTGALGQVVAEHLVREWGVRHLLLASRRG 3590

Qy   1303 SAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALFTRLGM-DSLMAVELRNRIEASLKL 1361
Db   3591 PEAPG--------------SDELASKLTGLGAEVTIVAADVSDPASVVELVGKTDPSHPL 3636

Qy   1362 KLSTTFLSTSPNIALLAQNLLDALATALS---LERV------AAENL-------RAGVQN 1405
Db   3637 ----------TGVVHAAGVLEDGVVTAQTPEGLARVWAAKAAAAANLHEATREMRLGLFV 3686

Qy   1406 DFVSSGA 1412
Db   3687 VFSSAAA 3693
```

RESULT 6
AAE10145

```
ID   AAE10145 standard; Protein; 5435 AA.
XX
AC   AAE10145;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysJ.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic.
XX
OS   Streptomyces noursei.
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17186.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 260-263; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS encoding Streptomyces noursei nystatin gene, NysJ.
XX
SQ   Sequence   5435 AA;

  Query Match             32.7%;  Score 2355.5;  DB 22;  Length 5435;
  Best Local Similarity   34.2%;  Pred. No. 5.6e-178;
```

```
Matches 655; Conservative 183; Mismatches 513; Indels 563; Gaps 47;

Qy          3 DRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPD 61
Db         32 NRMLVAAAKEPIAVVGMACRFPGAVDSPEALWEMVATGTDVISGFPDDRGWDLEALRNSG 91

Qy         62 PDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSA 121
Db         92 TDARDTDVSQRGGFLDCIADFDPGFFGISPREAVTMDPQQRLLLTTAWEAVERAGIDATT 151

Qy        122 LVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAV 181
Db        152 LRATRTGAFIGTNGQDYAYLLVRSLDDATGDV--GTGIAASAASGRLSYTLGLEGPALTV 209

Qy        182 DTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSA 241
Db        210 DTACSSSLVALHLAVQALRNGECGMALAGGVNVMATPGSLVEFSRQGGLARDGRCKAFAD 269

Qy        242 EADGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLK 301
Db        270 AADGTGWSEGAGVLLLERLSDAQRNGHPVLAVVRGSAVNQDGASNGFTAPNGPSQQRVIR 329

Qy        302 RALADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHP 361
Db        330 QALANAGLATGDIDAVEAHGTGTPLGDPIEAQSILATY--GQDRAHPVLLGSIKSNMGHT 387

Qy        362 EYASGITGLLKVVLSLQHGQIPAHLHAQALNPRISW--GDLRLTVTRARTPWPDWNTPRR 419
Db        388 QAASGVAGVIKMIMAMRHGVLPRTLHVDRPSTHVDWTTGSVEL-LTDAH-PWPETGRPRR 445

Qy        420 AGVSSFGMSGTNAHVVLEEAP-----AATCTPPAPERPAELLVLSARTASALDAQAARLR 474
Db        446 TGISSFGVSGTNAHVIVEQAPDTPAEAADDTPPRTPRTLPWL-LSARTGAALRDQATALL 504

Qy        475 DHLETYPSQCLG----DVAFSLATTRSAMEHRLAVAATSREGL--RAALDAAAQGQTSPG 528
Db        505 DHLDR-PDGDRGPTALDTAFSLATTRAALEHRLAV-VTGTDGTAGRDALTAWLAHGTAPD 562

Qy        529 AVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREV 588
Db        563 AHEGHA-AGRTRCAALFSGQGAQRLGMGRELHARFPVFARALDTAVDLLDAELGGTLREV 621

Qy        589 MWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGV 648
Db        622 IW----GTDDAPLNETGFTQPALFAVEVALYRLIESWGVAPDFVAGHSIGEIAAAHVAGV 677

Qy        649 FSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVI 708
Db        678 FSLEDACTLVAARAGLMQALPRGGAMVAVEATEDEVSPLLTD---GVAIAAINGPTSLVV 734

Qy        709 AGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVS 768
Db        735 SGDETATLAVAARLAEQGRRTTRLRVSHAFHSPLMDPMLAEFRAVAEGLSYGEPQIPVVS 794

Qy        769 NLSGKACTDE-VSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACM 827
Db        795 NLTGAVADGTLLGTADYWVRHVREAVRFADGIRALTDAGVGAFLELGPDGTLAALAQQSA 854

Qy        828 PDA--RPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFP-SGGRRVPLPTYPWQR 884
Db        855 PDAVSVPVL----RKDRDEEPAAVAALARLHTAGVPVDWTAFYAGTGAHRTDLPTYAFQY 910

Qy        885 ERYWID------------TKADDAARGD------------------------------ 900
Db        911 ERYWPKATYRPADATGLGLTAADHPLLGAAMSVAGSDELLLTGTLSLATHPWLADHVVGG 970

Qy        901 -------------RRAPGAGHDEVEE-------------------------------- 913
Db        971 MVFFPGTGFLELAVRAADQVGCDRVEELMLAAPLILPATGTVQMQIAVGAADDDGGRDLR 1030

Qy        914 -----------------------------------------------------------G 914
Db       1031 FFTRPGDDPDAAWAQHATGRITEGERVLALDTTTWPPRDAEPVDIDGLYDRYRANGLDYG 1090

Qy        915 GAVRG------------------------------------------GDRRS--- 924
Db       1091 PVFRGLRAVWRRDTEIYAEVALPEGTADADAFGLHPALFDAVLHSTLFASADGDDRSLLP 1150

Qy        925 -------------------------------------------------ARLDH 929
Db       1151 FAWNGVSLHAAGADALRVRITSCGPDAVEITAVDPQGRPVVSVESLTLRAAGPDAGTADH 1210

Qy        930 ------------------------------------------PPPESGRREKVEAAG- 944
Db       1211 RADAGSLFRMDWTPRTVHAPATPATWAVLGTDPIGLTEALTAAGPDTVTGLRDGVDALGE 1270

Qy        945 -----DRPE-------------------------------------------------- 948
Db       1271 LTAGDDRPVPDVVAVPLRGATDHGPAGAHDLTRTVLALLQEWLAEERFARSRLLLVTRGA 1330

Qy        949 -------------------------------------------------RLEIDEP-- 955
Db       1331 VADGERGPLDLAAAPVWGLVRSAQSENPGRLLLVDLDDTAESAAQLPLLPALLDADEPQA 1390

Qy        956 ----------------------------------GVLDHL-VLRVTERRAPGLG-EVE 977
Db       1391 VVREGTVRVGRLARLDSGRGLVPPPGTPWRLGSRAKGSLDGLALLPHPEARRPLTGHEVR 1450

Qy        978 IAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIA 1037
Db       1451 VGIRAAGLNFRDVLNALGMYPGDAG-------LFGSEAAGVVVEVGPEVTGLAPGDRVMG 1503

Qy       1038 LSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAA 1097
Db       1504 MLFGGFGPLGIADARLLTPVPADWSWETGASVPLVFLTAYYALKELGGLRAGEKVLVHAG 1563

Qy       1098 TGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVR--YVSDSRSDRFVADVRAWTGG 1155
Db       1564 AGGVGMAAIQIARHVGAEVFATA-SEGKWDVLRSLGVADDHIASSRTLDFEAAFAEVAGD 1622

Qy       1156 EGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRG 1215
Db       1623 RGLDVVLNALSGEFVDASMRLLGDGGRFLEMGKTDIRAADSVP-----DGLSYHSFDL-G 1676

Qy       1216 MMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTLG- 1274
Db       1677 MV--DPEHIQRMLLDLVELFDRGALAALPVRSWDVRRAGEAFRFMSLAQHIGKIVLTVPQ 1734

Qy       1275 --DPEVQIRIPTHAG--AG-------PSTGDRDLL--DRLASAAPAARA---------- 1310
Db       1735 PLDPDGTVLLTGGTGGLAGLLARHLVTEHGARHLLLAGRRGPDAPGAAALHAELTALGAE 1794

Qy       1311 ---AALEAFLRTQVSQVLRTPEIK----------VGAEALFTRLGMDSLMAVELRNRIE 1356
Db       1795 VTVAACDVADRTALAALLATVPAEHPLTAVVHTAGVLDDGTLTALNPDRLATV-LRPKVD 1853

Qy       1357 AS---------LKLKLSTTFLSTS-----PNIALLA--QNLLDALAT---ALSL 1391
Db       1854 AAWHLHDLTRHLDLAAFVLYSSTAGVMGGPGQANYAAGNTFLDALAAHRHALGL 1907
```

RESULT 7
ABG99872
ID   ABG99872 standard; Protein; 4106 AA.
XX
AC   ABG99872;
XX
DT   16-JAN-2003 (first entry)
XX
DE   S. cinnamonensis MonAV/polyketide synthase multi-enzyme MONS5.
XX
KW   Monensin; gene cluster; polyketide synthase; antibiotic;
KW   antihelminthic; insecticide; immunosuppressant; antifungal;
KW   antibacterial; polyether; mon BI; mon BII; mon CI; mon CII; mon H;
KW   mon RI; mon RII; mon T; mon AIX; mon AX.
XX

```
OS   Streptomyces cinnamonensis.
XX
PN   WO200168867-A1.
XX
PD   20-SEP-2001.
XX
PF   30-MAY-2000; 2000WO-GB02072.
XX
PR   28-MAY-1999;   99GB-0012563.
XX
PA   (BIOT-) BIOTICA TECHNOLOGY LTD.
XX
PI   Leadlay PF,  Staunton J,  Oliynyk M;
XX
DR   WPI; 2001-611393/70.
DR   N-PSDB; ABX04971.
XX
PT   New DNA sequence encoding polyketide synthase, useful for the
PT   production of polyketides such as antibiotic monensin -
XX
PS   Claim 6; Page 86-89; 212pp; English.
XX
CC   The invention relates to a DNA sequence which is a fully defined sequence
CC   of 103551 base pairs appearing as ABX04971, or its variant, that it is
CC   not a sequence encoding all or part amino acids 1-920 encoded by mon AI
CC   as given in the specification. The DNA is the S. cinnamonensis
CC   polyketide antibiotic monensin biosynthetic gene cluster. Also included
CC   are a recombinant cloning or expression vector comprising the gene
CC   cluster, a transformed host cell which has been transformed to
CC   contain the gene cluster (and is capable of expressing a corresponding
CC   polypeptide), a hybridization probe derived from the gene cluster
CC   (for identification and isolation of the same or analogous gene cluster,
CC   e.g. one which binds specifically to a region of the monensin gene
CC   cluster selected from mon BI, mon BII, mon CI, mon CII, mon H, Mon RI,
CC   mon RII, mon T, mon AIX and mon AX), the use of the mon RI gene or
CC   variant and a monensin promoter to control expression of a heterologous
CC   gene in Streptomyces cinnamonensis, a polypeptide encoded by a portion of
CC   the monensin gene cluster (preferably comprising mon BI, mon BII,
CC   mon AIX or mon AX or their mutants, alleles or variants), an epoxidase
CC   enzyme encoded by mon CI, a cyclase enzyme encoded by mon CII, producing
CC   S. cinnamonensis capable of enhanced levels of production of monensin
CC   comprising engineering it to overexpress the mon RI gene, S.
CC   cinnamonensis containing multiple copies of the mon RI gene and/or its
CC   variants, expressing a gene heterologous to S. cinnamonensis comprising
CC   transforming S. cinnamonensis with DNA encoding a heterologous gene and
CC   expressing the gene under control of the activator gene mon RI or
CC   actII/orf4 and 13-propyl erythromycin A. The processes and materials
CC   (enzyme systems, nucleic acids and vectors) are useful for preparing
CC   polyketides by recombinant synthesis. The polyketides are useful as
CC   insecticides, antibiotics, anthelminthics, antifungals, antibacterials or
CC   other pharmaceuticals. In particular the gene is useful for the
CC   production of monensin, an antibiotic polyether polyketide. The
CC   present sequence represents a protein encoded by the monensin gene
CC   cluster.
XX
SQ   Sequence   4106 AA;

  Query Match             32.6%;  Score 2351;  DB 22;  Length 4106;
  Best Local Similarity   32.4%;  Pred. No. 8.3e-178;
  Matches  608;  Conservative  185;  Mismatches  473;  Indels  608;  Gaps   35;

Qy          5 PIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPD 63
                |   | ::|:|||  :|: |||       | | |:     | ||  |   |  ||     ||||
Db       1850 PSAAATDEPVAIVAMACKYPGGATSPEGLWDLVAEGVDAVGAFPTGRGWDLERLFHPDPD 1909

Qy         64 APGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALV 123
                 || :      :|| |    |||:||||:||||| |||   |||||  || || | |  |: |
Db       1910 HPGTSYADEGAFLPDAGDFDAAFFGINPREALAMDPQQRLLLEASWEVLERAGIDPTTLK 1969

Qy        124 GTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDT 183
                || ||  ::|:    :| |  |      |:::  :  |        || :||::|  ||| || |||
Db       1970 GTPTGTYVGVMYHDYAAGLAQ---DAQLEGYSMLAGSGSVVSGRVAYTLGLEGPAVTVDT 2026

Qy        184 AYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEA 243
                | |||||::|||  |:||  |||:  ||||||::|  :|            |:  |  || |||||  |:| |
Db       2027 ACSSSLVSIHLAAQALRQGECTLALAGGVTVMATPEVFTGFSRQRGLAPDGRCKPFAAAA 2086

Qy        244 DGFGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRA 303
                ||  |   |||    |::|:||| ||     |  :|   |:||||:|  ||||:|||  |||  |||  |:::|
Db       2087 DGTGWGEGVGVLLLERLSDARRHGRRVLGVVRGSAVNQDGASNGLTAPNGPSQERVIRQA 2146

Qy        304 LADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEY 363
                ||   |  ::    |    ||  |||||||||||  |||  |  || ||   ||  :||||:|:|| :
Db       2147 LASGGLSSVDVDVVEGHGTGTTLGDPIEAQALLATYGQGRPVDRPLWLGSVKSNIGHTQA 2206

Qy        364 ASGITGLLKVVLSLQHGQIPAHLHAQALNPRISW--GDLRLTVTRARTPWPD-WNTPRRA 420
                |:|:  |::|:|::::|| :||  ||          :|  :  |    |  :||  |         |||:        ||||
Db       2207 AAGVAGVIKMVMAMRHGVVPASLHVDVPSPHVEWDSGAVRLAV--ESVPWPEVEGRPRRA 2264

Qy        421 GVSSFGMSGTNAHVVLEEAP------AATCTPPAPERPAE----LLVLSARTASALDAQA 470
                ||||||  |||||||::|   |              :  :  :  |||      :           |:|||:  ||  ||
Db       2265 GVSSFGASGTNAHVIVESVPDGLGEDSVSVSGEAPETETDGRLVPWVVSARSPQALRDQA 2324

Qy        471 ARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAV 530
                  ||||  :      :   :  ||  :||   ||:    |  |    |          |   |   :  |        ||  |:   |
Db       2325 LRLRDAVAADSTVSVQDVGWSLLKTRALFEQRAVVVGRERAELLSGLAVLAAGEEHPAVT 2384

Qy        531 RSIAD--SSRGKLAFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRP---- 584
                 ||    |  ::  |  :  :||:|||:|  :|||  |||:  :     ||||      |        |  :
Db       2385 RSREDGVAASGAVVWLFSGQGSQLVGMGAGLYERFPVFAAAFDEVCGL----LEGPLGVE 2440

Qy        585 ---LREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELV 641
                  ||||::  |           ||  |   :   |   ||    :    ||  ||  ||  ||  |:::|  ||||||:
Db       2441 AGGLREVVFRGPRE----RLDHTMWAQAGLFALQVGLARLWESVGVRPDVVLGHSIGEIA 2496

Qy        642 AACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVN 701
                ||  |||||  |  ||    :|  ||  |||    ||  ||||  :::|   |::||  |         :  ||:||||
Db       2497 AAHVAGVFDLADACRVVGARARLMGGLPEGGAMCAVQATPAELAADV--DDSGVSVAAVN 2554

Qy        702 APDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRR 761
                ||   ||:|      |   ||         ||  :||||  ||||||  ||  |||    |               |  :  |
Db       2555 TPDSTVISGPSGEVDRIAGVWRERGRKTKALSVSHAFHSALMEPMLAEFTEAIREVKFTR 2614

Qy        762 PSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLG 821
                |  :  |:||:||      ||  :|:||  ||  ||  |  |  |    |:  |  |    |:  ||  |||:||
Db       2615 PKVSLISNVSGLEAGEEIASPEYWARHVRQTVLFQPGI-AQVASTAGVFVELGPGPVLTT 2673

Qy        822 LVPACMPDA------RPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRV 875
                      :  |            |  |::|      |  |   :  :||:    |        |     |  |:  ||       ||
Db       2674 AAQHTLDDVTDRHGPEPVLVSSLAGERPEESAFVEAMARLHTAGVAVDWSVLF--AGDRV 2731

Qy        876 P----LPTYPWQRERYWIDTK-------------------------------------- 892
                |      ||||  :||||:|:    :
Db       2732 PGLVELPTYAFQRERFWLSGRSGGGDAATLGLVAAGHPLLGAAVEFADRGGCLLTGRLSR 2791

Qy        893 ------AD---------------------------------------------------- 894
                       ||
Db       2792 SGVSWLADHVVAGAVLVPGAALVEWALRAGDEVGCVTVEELMLQAPLVVPEASGLRVQVV 2851

Qy        895 --------------------DAARGD---------------------------------- 900
                                    ||  ||
Db       2852 VEEAGEDGRRGVQIYSRPDADAVSGDDSWICHATGTLTPQHTDAPNDGLAGAWPAAGAVP 2911

Qy        901 ------------------------------------------------------------ 900

Db       2912 VDLAGFYERVADAGYAYGPGFQGLRAVWRHGQDLLAEVVLPEAAGAHDGYGIHPALLDAT 2971

Qy        901 -------------------------------RRAPGAG----------HDEVEE------- 913
                                               || ||                    ||| |
Db       2972 LHPALLLDWPGEVQDDDGKVWLPFTWNQVSLRAAGAATVRVRLSPGEHDEAEREVQVLVA 3031

Qy        914 ------------------------------------------------------------ 913

Db       3032 DATGTDVLSVGSVTLRPADIRQLQAVPGHDDGLFSVDWTPLPLSRTDVSQTDADGDADWV 3091

Qy        914 -------------------------------GGAVRGGDR------------------- 922
```

```
                                                   || : | ||
Db       3092 VLSDGVGSLADVVSAAGGEAPWAVVAPVGASAGGGLAGFDRREGLDGRLVVERVLSLVQE 3151

Qy        923 ----------------------------------------RSARLDHP----------- 930
                                                      |||: ::|
Db       3152 FLAAPELAESRLLVLTRGAVATGGDGDGDVDASAAAVWGLVRSAQSENPGRFILLDVDMD 3211

Qy        931 ----------------------------PPESGRR------------------------ 937
                                          | :|||
Db       3212 VDVDVDMDVDVDVDVDVDVDGDGNGSDLDPDLNGRRLPHATLRHAAEELDEPQLALRDGQ 3271

Qy        938 ----EKVEAAG--------DRPFRLE------------IDEPGVLDHLVLRVTERRAPGL 973
                  | | |      . || :||:           :  |||:: |          |
Db       3272 LLVPRLVRATGGGLVVAPTDRAWRLDKGSAETLESVAPVAYPGVMEPL----------GP 3321

Qy        974 GEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQ 1033
              |:| : : |||::| || ::|||||   :           |||| || :   |  |  | ||
Db       3322 GQVRLGIHAAGINFRDVLVSLGMVPGQVG--------LGGEGAGVVTETGPDVTHLSVGD 3373

Qy       1034 PVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQPGERVL 1093
               |: :  |:|          :| | ||    :|||||||||||| |   :| |: |||||
Db       3374 RVMGVLHGSFGPTAVADTRMVAPVPQGWDMRQAAAMPVAYLTAWYGLVELAGLKAGERVL 3433

Qy       1094 IHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGV--RYVSDSRSDRFVADVRA 1151
              ||||||||:|||| |:|:|||| |||  :  |    || :|:   : : ||    |    |
Db       3434 IHAATGGVGMAAVQIARHLGAEVFATA-SAAKHVVLEEMGIDAAHRASSRDLAFEDTFRQ 3492

Qy       1152 WTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLR-PFLRNLSFSL 1210
               | | |:|||||||:|| || |  ||   |||:|:|| |        ::     |  :  : |
Db       3493 ATDGRGMDVVLNSLTGEFIDASLRLLGDGGRFLEMGKTDVRTPEEVAAEYPGVTYTVYDL 3552

Qy       1211 VDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQAQHLGKLV 1270
               |  |     | |:  :: ||    |:|   | |: : |: :  :||| |:||:| ||||
Db       3553 VTDAG-----PDRIAVMMSELGERFASGALDPLPVRSWPLDKAREAFRFMSQAKHTGKLV 3607

Qy       1271 LTLG---DPEVQIRIPTHAGA--------------------------GPSTGDRDLLDR 1300
              | :    ||:  : |     ||                     .     | :|  :| ||
Db       3608 LDVPAPLDPDGTVLITGGTGALGQVVAEHLVREWGVRHLLLASRRGLDAPGSG--ELADR 3665

Qy       1301 LASAAPAARAAALE 1314
              |:         || :
Db       3666 LSDLGAEVTVAAAD 3679
```

```
RESULT 8
AAE10129
ID   AAE10129 standard; Protein; 11096 AA.
XX
AC   AAE10129;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysC.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          35..455
FT                   /label= KS3_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          546..858
FT                   /label= AT3_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          872..1073
FT                   /label= DH3_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1381..1628
FT                   /label= KR3_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          1662..1735
FT                   /label= ACP3_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          1757..2180
FT                   /label= KS4_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2291..2603
FT                   /label= AT4_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2617..2818
FT                   /label= DH4_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          3124..3371
FT                   /label= KR4_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3407..3480
FT                   /label= ACP4_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3501..3924
FT                   /label= KS5_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          4032..4346
FT                   /label= AT5_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4360..4561
FT                   /label= DH5_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4953..5239
FT                   /label= ER5_domain
FT                   /note= "Enoylreductase (ER) domain"
FT   Domain          5248..5495
FT                   /label= KR5_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          5528..5601
FT                   /label= ACP5_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          5623..6046
FT                   /label= KS6_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          6166..6478
FT                   /label= AT6_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          6492..6704
FT                   /label= DH6_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          7038..7281
FT                   /label= KR6_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          7315..7388
FT                   /label= ACP6_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          7408..7831
FT                   /label= KS7_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          7939..8253
FT                   /label= AT7_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          8267..8470
FT                   /label= DH7_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          8812..9086
FT                   /label= KR7_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          9120..9193
FT                   /label= ACP7_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          9214..9637
FT                   /label= KS8_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          9758..10072
FT                   /label= AT8_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          10086..10289
FT                   /label= DH8_domain
```

```
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          10657..10904
FT                   /label= KR8_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          10939..11012
FT                   /label= ACP8_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17184.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 170-176; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei nystatin
CC   gene, NysC.
XX
SQ   Sequence   11096 AA;

  Query Match             32.3%;  Score 2331.5;  DB 22;  Length 11096;
  Best Local Similarity   32.9%;  Pred. No. 1.4e-175;
  Matches  629;  Conservative  182;  Mismatches  519;  Indels  583;  Gaps   34;

Qy         10 AEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGKT 68
              |:|||  :|| :|| ||||          | |::     |:     |  | ||     :||||:  | :
Db       3499 ADDPIVVVGMACRFPGGVTSPEDLWRLVDDGTDAITTFPTNRGWDLDNLYDPDPEHFGTS 3558

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETG 128
                       || :    || :|||:|||||   |   |||||  |||:| | | |   | |: ||
Db       3559 YTRSGGFLHEAGEFDPAFFGMSPREALATDSQQRLLLESSWEAIERAGIDPLTLRGSATG 3618

Qy        129 VFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSS 188
              || |:  |:| : |            | :     | |:  || :||:|||||  || | |||| |||
Db       3619 VFAGVMYSDYGSIL----GGKEFEGFQGQGSAGSVASGRVSYALGFEGPAVTVDTACSSS 3674

Qy        189 LVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR 248
              |||:| | |:||:||||  ||||||::| :|||  |   |: | || ||| |||:   ||| |
Db       3675 LVALHWAAQALRAGECSLALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKAFAEAADGVGW 3734

Qy        249 GEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAG 308
               ||  ::||:|  |  |   :|  |||||||||:| ||||:|||  ||| ||: |:::||| |
Db       3735 SEGVGILVLERQSDAVRNGHEILAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGG 3794

Qy        309 CAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
               : :  |  |||||||||||||| |||  |  ||   ||    || :||:|:|:|| : |:|:
Db       3795 LSTADVDAVEAHGTGTTLGDPIEAQALLATYGRDRDPENPLWLGSLKSNIGHTQAAAGVA 3854

Qy        369 GLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMS 428
              |::|:|::::|| :|  ||  |  :   :  :   :  :   :| ||:   ||||||||:|
Db       3855 GVIKMVMAMRHGVLPQTLHVDAPSSHVDWSVGAVELLTEQTAWPETGRVRRAGVSSFGIS 3914

Qy        429 GTNAHVVLE-----EAPAATCTPPAPERPAEL-LVLSARTASALDAQAARLRDHLETYPS 482
              ||||||::|      |:|||   |:   |  :  | || ::  ||  |||||  |     |:
Db       3915 GTNAHVIVEQPALVESPAA---EPSGREPGVVPLPLSGKSPEALRDQAARLLAGLAERPA 3971

Qy        483 QCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLA 542
                    |: :||||||| :||  | || |    || | |        ||  : |:  |: |
Db       3972 LRPLDLGYSLATTRSAFDHRAVVLATDRADAVRALTALAAADADLSAV--VGDTRTGRHA 4029

Qy        543 FLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQEL--DRPLREVMWAEPASVDAAL 600
               ||:|||:|·||||  ||: :  |  ||  |: :   :   |    |||||  :    |  |
Db       4030 VLFSGQGSQRLGMGRELYERFPVFAEALDVAIDHLDAALPAQASLREVMWGD----DVEL 4085

Qy        601 LDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAA 660
              ||:| :|||||  | ||  |  |||.|: |||||||: || | |||||||  |||
Db       4086 LDETGWTQPALFAVEVALFRLVESWGVRPDFVAGHSIGEIAAAHVVGVFSLEDACRLVAA 4145

Qy        661 RGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAA 720
               |  ||||||  ||||::|:|  |  |  :|          ||||||  |  ||::||       :|
Db       4146 RATLMQALPTGGAMIAIQAAEDEVTQHLTD---DVSIAAVNGPTSVVVSGAESAARTVAD 4202

Qy        721 AMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACT-DEV 779
               :|   |  :|   |  ||||||||||  |||   |   |||  :||  |::  :||||:|:   |  |::
Db       4203 RLAENGRKTTRLRVSHAFHSPLMDPMLAEFRAVAEGLSYATPTLPVVSNLTGRLATADDL 4262

Qy        780 SSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMP-DARPALLASS 838
               |  ||  ||  ||  |||||||  |    |    |||:|:||   |   :   :  ||  |  : :
Db       4263 CSAEYWARHVREAVRFADGVSTLENEGVTTFLELGPDGVLSAMAQQSLTGDA--ATVPAL 4320

Qy        839 RAGRDEPATVLEALGGLWAVGGLVSWAGLFP-SGGRRVPLPTYPWQRERYW----IDTK- 892
               |  |||   :  | ||    |      |   |  |  ||  ||  ||||:|  ||    : |
Db       4321 RKDRDEETSALTALAHLHTAGLRVDWAAFFAGSGATRVDLPTYAFQHATYWPTGTLPTAH 4380

Qy        893 --------------------------------------ADDAARGDRRAPG------- 905
                                                      ||  |    |    ||
Db       4381 AAAVGLTAAEHPLLNGSVELAEGEGVLFTGRLSLQSHPWLADHAVMGQVLLPGTALLELA 4440

Qy        906 ------AGHDEVE-----------EGGAV----------------------------- 917
                     || |  ||              | |||
Db       4441 FRAGDEAGCDRVEELTLAAPLVLPERGAVQTQVRVGVADDTGRRTVTVHSRPEHATDVSW 4500

Qy        918 -----------------------------------------------------------R 918
                                                                          |
Db       4501 TQHATGTLTMGSAPADTGFDATAWPPADAEPLATDDCYARFTTLGFAYGPVFQGLRAAWR 4560

Qy        919 GGD--------------------------------------------------------- 921
               ||
Db       4561 AGDVLYAEVALAESTGDEATAFGLHPALLDAALHASLVAHEGEESNGGLPFSWEGATLYA 4620

Qy        922 ------------------------------------------------------------ 921

Db       4621 TGATALRVRLTPTGTDGRSVAIAVADTAGRPVAAIDNLVSRRVSGDQLTGAAGLARDALF 4680

Qy        922 ------------------------------------------------------------ 921

Db       4681 TLDWNPVPENLVPENPVPENTGGGHAQDQDGRPAAATVALVGADGTAIAADLTAAGIHTT 4740

Qy        922 ------------------------------------------------------------ 921

Db       4741 LHPDLTTLATTDADVPKTVLIPLTGTGTGTGTGTESTDGIGTGAAESDASAPSPAEVAHT 4800

Qy        922 ----------------------------------------------------RRSAR------ 926
                                                                  |||:
Db       4801 LSTAALALVQEWTAQERFAGSRLAFVTTGATAAGGTDVMDVAAAAVWGLVRSAQSEAPDT 4860

Qy        927 ---LDHPPPESGRREKVEAA---------------------------------------- 943
```

```
Db       4861 FVLIDRDPGPAGTHDRTAAAERGQLLLRALHTDEPQLALRDGGVLAARLARFDTAAALTP 4920

Qy        944 -GDRPFRLEIDEPGVLDHLVLR-VTERRAPGLG-EVEIAVDAAGLSFNDVQLALGMVPDD 1000

Db       4921 PADRAWRLDSTAKGSLNGLALTPYPAALAPLTGHEVRVEVRAAGLNFRDVLNALGMYPGD 4980

Qy       1001 LPGKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQA 1060

Db       4981 DVGS------FGSEAAGVVVEVGPEVTGLAPGDQVMGMITGSFGSLAVDDARRLARLPED 5034

Qy       1061 LSAIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATA 1120

Db       5035 WSWETGASVPLVFLTAYYALKELGGLRAGEKVLVHAGAGGVGMAAIQIARHVGAEVFATA 5094

Qy       1121 GTPEKRAYLESLGVR--YVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLR 1178

Db       5095 -SEGKWDVLRSLGVADDHIASSRTLDFEAAFAEVAGDRGLDVVLNSLAGDFVDASMRLLG 5153

Qy       1179 SHGRFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAG 1238

Db       5154 DGGRFLEMGKTDIRAADSVP-----DGLSYQSFDLAWVV---PETIGTMLAELMDLFRTG 5205

Qy       1239 VFTPPPIATLPIARVADAFRSMAQAQHLGKLVLTL---GDPEVQIRIPTHAGAG------ 1289

Db       5206 ALRPLPVRTWDVRHAKDAFRFMSMAKHIGKIVLTLPRSWKPEGTVLVT--GGTGGLGGLV 5263

Qy       1290 -----PSTGDRDLLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAEALFTRLGM- 1343

Db       5264 ARHLVRSCGVRHLLLTSRSGVGAAGAAGLVAELES------------LGARVVVAACDVG 5311

Qy       1344 DSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATALSLERVAA 1396

Db       5312 DGSAVAELVAGVSESYPLSA----------VVHAAGVLDDGVVGSLTPERLAA 5354
```

RESULT 9
ABP57679

```
ID   ABP57679 standard; Protein; 2149 AA.
XX
AC   ABP57679;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Saccharopolyspora busB butenyl-spinosyn biosynthetic gene product.
XX
KW   Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;
KW   metabolite; spinosyn.
XX
OS   Saccharopolyspora sp..
XX
PN   WO200279477-A2.
XX
PD   10-OCT-2002.
XX
PF   28-MAR-2002; 2002WO-US09968.
XX
PR   30-MAR-2001; 2001US-280175P.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Hahn DR,  Jackson JD,  Bullard BS,  Gustafson GD,  Waldron C;
PI   Mitchell JC;
XX
DR   WPI; 2003-058434/05.
DR   N-PSDB; ABV75557.
XX
PT   New butenyl-spinosyn biosynthetic genes, useful for increasing the
PT   production of butenyl-spinosyn insecticidal macrolides, or for changing
PT   the metabolites or products produced by spinosyn-producing
PT   microorganisms  -
XX
PS   Claim 1; Page 134-141; 218pp; English.
XX
CC   The invention relates to a novel DNA molecule comprising a DNA sequence
CC   that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
CC   PKS (polyketide synthetase) domain, or a spinosyn PKS module. The
CC   butenyl-spinosyn biosynthetic genes are useful for increasing the
CC   production of butenyl-spinosyn insecticidal macrolides. The genes are
CC   also useful for changing the metabolites or products produced by
CC   spinosyn-producing microorganisms. The present sequence represents one of
CC   the butenyl-spinosyn biosynthetic polypeptides of the invention.
XX
SQ   Sequence   2149 AA;

  Query Match             32.2%;  Score 2319.5;  DB 24;  Length 2149;
  Best Local Similarity   32.8%;  Pred. No. 1e-175;
  Matches  610;  Conservative  198;  Mismatches  547;  Indels  507;  Gaps   32;

Qy          9 AAE--DPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAP 65

Db         30 AAEKGDPIAIVAMSCRYPGQVSSPEDLWQLAAGGVDAISEVPGDRGWDLAGVFDPDSDRP 89

Qy         66 GKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGT 125

Db         90 GTSYACAGGFLQGVSEFDAGFFGISPREALAMDPQQRLLLEVAWEVFERAGLEQRSTRGS 149

Qy        126 ETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAY 185

Db        150 RVGVFVGTNGQDYASWL--RTPPSEVAGHVLTGGAAAILSGRVAYSFGFEGPAVTVDTAC 207

Qy        186 SSSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADG 245

Db        208 SSSLVALHLAGQALRAGECDLALAGGVTVMSTPKAFLEFSRQRGLAADGRCKSFAAAADG 267

Qy        246 FGRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALA 305

Db        268 TGWGEGAGLLLLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGSSQARVITQALA 327

Qy        306 DAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYAS 365

Db        328 SAGLSVSDVDAVEAHGTGTRLGDPIEAQALIATYGRDRDPARPLWLGSVKSNIGHTQAAA 387

Qy        366 GITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSF 425

Db        388 GVAGVIKMVMAMRHGQLPRTLHVDAPSPEVDWSAGTVQLLTENMLWPESGRVRRAGVSSF 447

Qy        426 GMSGTNAHVVLEEAPAATCTPPAPER------PAELLVLSARTASALDAQAARLRDHLET 479

Db        448 GISGTNAHVILEQPTGETRQSAGPDSGSVVDVPVPVPWMVSGKTPDALGAQADTLMSYLDD 507

Qy        480 YPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRG 539

Db        508 RVDVPSLDIAYSLAMTRSALDERAVVLGPDRETLLSGLKALSAGHEASGVVTGSVGTG-G 566

Qy        540 KLAFLFTGQGAQTLGMGRGLYDVWSAFREAFD-LCVRL---FNQELDRPLREVMWAEPAS 595

Db        567 RIGFVFSGQGGQWLGMGRGLYRAFPVFAAAFDEACAELEAHLGQEVG--VRDVVFGS--- 621

Qy        596 VDAALLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAV 655

Db        622 -DAQLLNRTLWAQSGLFALQVGLLKLLDSWGVRPSAVLGHSVGELAAAFAAGVLSLSDAA 680

Qy        656 FLVAARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPV 715

Db        681 RLVAGRARLMQALPSGGGMLAVAAGEEQLRPLLADHGDRVGLAAVNVAESVVLSGDRDVL 740

Qy        716 HAIAAAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKA- 774

Db        741 DDIAGRLDGQGVRTRWLRVSHAFHSYRMDPMLDEFAEIARAVDYRRCELPIVSTLTGKLD 800

Qy        775 CTDEVSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACM----PDA 830

Db        801 DAGRMSGPDYWVRQVREPVRFADGAQALVEHDVATIVEIGPDGALSALIQECVAASDQSR 860

Qy        831 RPALLASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFP-SGGRRVPLPTYPWQRERYWI 889
```

```
               | | : : |  |||   :  ||  :   || | |    |  :| ::| |||| :||:|||:
Db        861 RVAAVPAMRRNRDEAQNLTTALAQVHVRGGAVDWRSFFAGTGAKQVELPTYAFQRQRYWL 920

Qy        890 DTKADDAARGDRRA-----------PGAGHDEVEEGG---------------------- 915
              . :         |   |         | || |||   |
Db        921 EPSDSGDVTGAGLAGAEHPLLGAVVPVAGGDEVLLTGRISVGTHPWLAEHRVLGEVIVPG 980

Qy        916 ------AVRGGDR------RSARLDHP-----------------PPESGRREKV------ 940
                    |:  |:|        |: |                 |  ||||  |
Db        981 TALLEIALHAGERLGCERVEELTLEAPLVLPERGAMQVQLRVGAPENSGRRPMVLYSRPE 1040

Qy        941 --------------------EAAGD----------------------------------- 945
                                  |||||
Db       1041 GAADHDWTRHATGRLAPGGGEAAGDLADWPAPGALPVDLDEFYRDLAEHGLEYGPIFQGL 1100

Qy        946 ------------------------------------------------------------ 945

Db       1101 KAAWRQGDEVYAEAALPGTEDSGFGVHPALLDAALHATAVRDMDGAWLPFQWEGVCLHAR 1160

Qy        946 --------------------------RP-------------------------------- 947
                                        ||
Db       1161 AASALRVRVVPAGDDAKSLLVCDGTGRPVISVDRLVFRSAAAGRTGARRQAHRARLYRLG 1220

Qy        948 ------------------------------------------------------------ 947

Db       1221 WPTVQLPTSAQPPSCVLLGTSEVSSDMQVYPDLRSLTAALDAGAEPPGVVIAPTPPGGGQ 1280

Qy        948 ------------------------------------------------------------ 947

Db       1281 TADVRESTRHALDLVQGWLADQRLNDSRLFLVTRGAVAVEPGEPVTDLAQAALWGLLRST 1340

Qy        948 ------------------------------------------------------------ 947

Db       1341 QTEHPDRFVLVDVAEPAQLLPALPGVLACGEPQLALRRGGAHAPRLAGLGGDDVLPVPDS 1400

Qy        948 --FRLEIDEPGVLDHLVLRVTERRAP---GLGEVEIAVDAAGLSFNDVQLALGMVPDDLP 1002
                :|||  || || | | : |  |    | |:| ||: |||::| |   :|||| |
Db       1401 MGWRLEATSPGTLDGLAL-LDEPAATASLGDGQVRIAMRAAGVNFRDALIALGMYPG--- 1456

Qy       1003 GKPNPPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALS 1062
                      |||| || :|  | || ||  |   | |: :    ||           :|   |   |
Db       1457 -----AASLGGEGAGVVVETGPGVTGLAPGDRVMGMIPKAFGPLAVADHRMVTRIPAGWS 1511

Qy       1063 AIEAAAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGT 1122
                :||::|: :||:|||   :|  |:|||  :|:|:|  ||||:||:|  |:|:||||:|||
Db       1512 FAQAASVPIVFLTAYYALVDLAGLRPGESLLVHSAAGGVGMAAIQLARHLGAEVYATASE 1571

Qy       1123 PEKRAYLESLGVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGR 1182
                : :|     |    :: ||:  |           ||| |||||||||:|:  |  |  :|   ||
Db       1572 DKWQAV--ELTRERLASSRTCDFEKQFLGATGGRGVDVVLNSLAGDFADASLRMLPRGGR 1629

Qy       1183 FVELGKRDCYADNQLG-LRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFT 1241
              |:||||  |      ::        | :   :|   |:              | |:  :|:||: |  ||
Db       1630 FLELGKTDVRDPVEVADAHPGVSYQAFDTVE------AGPQRIGEMLDELVELFEGGVLE 1683

Qy       1242 PPPIATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHAGAGP-----STGDRD 1296
              | |:    : :  :| |  ::||:|:||||||:             |    ||         ||
Db       1684 PLPVTAWDVRQAPEALRHLSQARHVGKLVLTMP--------PAWDTAGTVLVTGGTG--- 1732

Qy       1297 LLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAE--ALFTRLGMD-SLMAVELRN 1353
                    |   |   ||    :|  :|  |       | |     |  |   | |  ||  |  :: :
Db       1733 -----ALGAEVARHLVIEHGVRNLVLVSRRGPAASGAAELVAQLTAYGAEVSLQACDVAD 1787

Qy       1354 RIEASLKLKLSTTFLSTSPNIALLAQNLLDALATALSLERV---------AAENLRAGVQ 1404
              | |   |:       |    :    |  | | :| :|: :|:              | ||     :
Db       1788 R-ETLAKVLAGIPDEHTLTAVVHAAGVLDDGVAESLTAQRLDHVLRPKVDGARNLHELIA 1846

Qy       1405 ND 1406
               |
Db       1847 PD 1848

RESULT 10
AAY39298
ID   AAY39298 standard; Protein; 2152 AA.
XX
AC   AAY39298;
XX
DT   01-DEC-1999  (first entry)
XX
DE   SpnB a polyketide synthase.
XX
KW   Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW   microlides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW   PKS; extender module; initiator module; acyl transferase domain; AT;
KW   acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW   dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW   insecticide.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..424
FT                   /label= KS2
FT                   /note= "Beta-ketosynthase domain: part of
FT                   extender module 2"
FT   Domain          536..886
FT                   /label= AT2
FT                   /note= "Acyl transferase domain: part of extender
FT                   module 2"
FT   Domain          892..1077
FT                   /label= ACP2
FT                   /note= "Acyl carrier protein domain: part of extender
FT                   module 2"
FT   Domain          1338..1683
FT                   /label= ER2
FT                   /note= "Enoyl reductase domain: part of extender
FT                   module 2"
FT   Domain          1687..1866
FT                   /label= KR2
FT                   /note= "BEta-ketoreductase domain: part of extender
FT                   module 2"
FT   Domain          1955..2034
FT                   /label= ACP2
FT                   /note= "Acyl carrier protein domain: part of extender
FT                   module 2"
XX
PN   WO9946387-A1.
XX
PD   16-SEP-1999.
XX
PF   16-FEB-1999;   99WO-US03212.
XX
PR   09-MAR-1998;   98US-0036987.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Baltz RH,  Broughton MC,  Crawford KP,  Madduri K,  Merlo DJ;
PI   Treadway PJ,  Turner JR,  Waldron C;
XX
DR   WPI; 1999-551414/46.
DR   N-PSDB; AAZ21501.
XX
PT   New spinosyn biosynthetic genes from Saccharopolyspora spinosa, useful
PT   for production of insecticidal spinosyn compounds  -
XX
PS   Claim 1; Page 96-103; 190pp; English.
XX
CC   This is the amino acid sequence of the product of the spnB gene. The
CC   protein is involved in spinosyn biosynthesis. The SpnB gene is one of 23
CC   genes and open reading frames contained in an 80kb DNA sequence
CC   AAZ21501. Spinosyns are insecticidal microlides which are useful for the
CC   control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC   occurs via stepwise condensation and modification of carboxylic acid
CC   precursors generating a linear polyketide which is modified further. The
CC   DNA sequence contains a central region of approximately 55kb which has
```

```
CC   homology to the DNA encoding the polyketide synthases (PKS) of known
CC   macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with
CC   stop codons at the end of acyl carrier protein (ACP) domains. Together
CC   the PKS polypeptides (AAY39297-Y39301) of which this sequence is one,
CC   form a complex consisting of an initiator module, spnA, and several
CC   extender modules spnB-spnE. Each extender module adds a specific acetyl
CC   Co-A precursor to a growing polyketide chain, and modifies the beta-keto
CC   group in a specific manner. A module in a PKS polypeptide consists of
CC   several domains with specific functions. The initiator module has an
CC   acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain.
CC   The extender modules have the same domains plus a beta-ketosynthase (KS)
CC   domain and optionally a beta-ketoreductase domain, a dehydratase (DH)
CC   domain, and an enoyl reductase (ER) domain. The last extender module
CC   terminates with a thiolester domain. The products of the genes present
CC   in the upstream region the PKS genes have been assigned names spnF-spnS
CC   AAY39302-Y39315 and are responsible for different modifications in
CC   spinosyn biosynthesis. There are also two ORFs ORFL15 and ORFL16 present
CC   immediately upstream of spnS, producing polypeptides AAY39316-Y39317,
CC   and two ORFs ORF1 and ORF2 present downstream of the PKS region
CC   producing polypeptides AAY39318-Y39319. The genes are useful to improve
CC   yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,
CC   or interruption of steps in spinosyn biosynthesis. The modified
CC   spinosyns may be a new insect control agent or serve as substrates for
CC   further chemical modification and the creation of new semi-synthetic
CC   spinosyns. The genes are also useful to isolate similar sequences from S.
CC   spinosa or other species by hybridization.
XX
SQ   Sequence   2152 AA;

  Query Match             31.5%;  Score 2273;  DB 20;  Length 2152;
  Best Local Similarity   32.7%;  Pred. No. 5.3e-172;
  Matches  600;  Conservative  199;  Mismatches  542;  Indels  496;  Gaps   32;

Qy     11 EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGKTP 69
Db     36 DDPIAIVAMSCRYPGQVSSPEDLWQLAAGGVDAISEVPGDRGWDLDGVFVPDSDRPGTSY 95

Qy     70 VTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGV 129
Db     96 ACAGGFLQGVSEFDAGFFGISPREALAMDPQQRLLLEVAWEVFERAGLEQRSTRGSRVGV 155

Qy    130 FIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSSL 189
Db    156 FVGTNGQDYASWL--RTPPPAVAGHVLTGGAAAVLSGRVAYSFGFEGPAVTVDTACSSSL 213

Qy    190 VAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGRG 249
Db    214 VALHLAGQALRAGECDLALAGGVTVMSTPKVFLEFSRQRGLAPDGRCKSFAAGADGTGWG 273

Qy    250 EGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAGC 309
Db    274 EGAGLLLLERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGSSQQRVITQALASAGL 333

Qy    310 AASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITG 369
Db    334 SVSDVDAVEAHGTGTRLGDPIEAQALIATYGRDRDPGRPLWLGSVKSNIGHTQAAAGVAG 393

Qy    370 LLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSG 429
Db    394 VIKMVMAMRHGQLPRTLHVESPSPEVDWSAGTVQLLTENTPWPRSGRVRRVGVSSFGISG 453

Qy    430 TNAHVVLEEAPAATCTPPAP------ERPAELLVLSARTASALDAQAARLRDHLETYPSQ 483
Db    454 TNAHVILEQPPGVPSQSAGPGSGSVVDVPVVPWMVSGKTPEALSAQATALMTYLDERPDV 513

Qy    484 CLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAF 543
Db    514 SSLDVGYSLALTRSALDERAVVLGSDRETLLCGVKALSAGHEASGLVTGSVGAG-GRIGF 572

Qy    544 LFTGQGAQTLGMGRGLYDVWSAFREAFD-LCVRL---FNQELDRPLREVMWAEPASVDAA 599
Db    573 VFSGQGGQWLGMGRGLYRAFPVFAAAFDEACAELDAHLGQEIG--VREVV----SGSDAQ 626

Qy    600 LLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVA 659
Db    627 LLDRTLWAQSGLFALQVGLLKLLDSWGVRPSVVLGHSVGELAAAFAAGVVSLSGAARLVA 686

Qy    660 ARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIA 719
Db    687 GRARLMQALPSGGGMLAVPAGEELLWSLLADQGDRVGIAAVNAAGSVVLSGDRDVLDDLA 746

Qy    720 AAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKA-CTDE 778
Db    747 GRLDGQGIRSRWLRVSHAFHSYRMDPMLAEFAELARTVDYRRCEVPIVSTLTGDLDDAGR 806

Qy    779 VSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDA----RPAL 834
Db    807 MSGPDYWVRQVREPVRFADGVQALVEHDVATVVELGPDGALSALIQECVAASDHAGRLSA 866

Qy    835 LASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFP-SGGRRVPLPTYPWQRERYW----- 888
Db    867 VPAMRRNQDEAQKVMTALAHVHVRGGAVDWRSFFAGTGAKQIELPTYAFQRQRYWLVPSD 926

Qy    889 ------------------------------------------------------------ 888
Db    927 SGDVTGAGLAGAEHPLLGAVVPVAGGDEVLLTGRISVRTHPWLAEHRVLGEVIVAGTALL 986

Qy    889 --------------------------------------------------IDTKADDAAR 898
Db    987 EIALHAGERLGCERVEELTLEAPLVLPERGAIQVQLRVGAPENSGRRPMALYSRPEGAAE 1046

Qy    899 GD-------RRAPGAGH------------------------------------------- 908
Db   1047 HDWTRHATGRLAPGRGEAAGDLADWPAPGALPVDLDEFYRDLAELGLEYGPIFQGLKAAW 1106

Qy    909 ---DEV---------EEGG----------------------------------------- 915
Db   1107 RQGDEVYAEAALPGTEDSGFGVHPALLDAALHATAVRDMDDARLPFQWEGVSLHAKAAPA 1166

Qy    916 -------------------------------------AVRGGDRRSA------RLDHP-- 930
Db   1167 LRVRVVPAGDDAKSLLVCDGTGRPVISVDRLVLRSAAARRTGARRQAHQARLYRLSWPTV 1226

Qy    931 --------------------------------------------------PPESGRREKV 940
Db   1227 QLPTSAQPPSCVLLGTSEVSADIQVYPDLRSLTAALDAGAEPPGVVIAPTPPGGGRTADV 1286

Qy    941 E----------------------------------------------------------- 941
Db   1287 RETTRHALDLVQGWLSDQRLNESRLLLVTQGAVAVEPGEPVTDLAQAALWGLLRSTQTEH 1346

Qy    942 -----------------------AAGD----------------------------RPFR 949
Db   1347 PDRFVLVDVPEPAQLLPALPGVLACGEPQLALRRGGAHAPRLAGLGSDDVLPVPDGTGWR 1406

Qy    950 LEIDEPGVLDHLVLRVTERRAP---GLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPN 1006
Db   1407 LEATRPGSLDGLAL-VDEPTATAPLGDGEVRIAMRAAGVNFRDALIALGMYPG------- 1458

Qy   1007 PPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEA 1066
Db   1459 -VASLGSEGAGVVVETGPGVTGLAPGDRVMGMIPKAFGPLAVADHRMVTRIPAGWSFARA 1517

Qy   1067 AAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKR 1126
Db   1518 ASVPIVFLTAYYALVDLAGLRPGESLLVHSAAGGVGMAAIQLARHLGAEVYATASEDKWQ 1577

Qy   1127 AYLESLGVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVEL 1186
Db   1578 AV--ELSREHLASSRTCDFEQQFLGATGGRGVDVVLNSLAGEFADASLRMLPRGGRFLEL 1635

Qy   1187 GKRDCYADNQLG-LRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPI 1245
Db   1636 GKTDVRDPVEVADAHPGVSYQAFDTVE------AGPQRIGEMLHELVELFEGRVLEPLPV 1689

Qy   1246 ATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHAGAGP-----STGDRDLLDR 1300
```

```
                 : :  :| |  ::||:|:||||||:          |    ||        ||
Db       1690 TAWDVRQAPEALRHLSQARHVGKLVLTMP--------PVWDAAGTVLVTGGTG------- 1734

Qy       1301 LASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAE--ALFTRLGMD-SLMAVELRNRIEA 1357
                |  |  ||    :|  :|   |     | |       ||   |   |  | : || |  :: :| |
Db       1735 -ALGAEVARHLVIERGVRNLVLVSRRGPAASGAAELVAQLTAYGAEVSLQACDVADR-ET 1792

Qy       1358 SLKLKLSTTFLSTSPNIALLAQNLLDALATALSLERV 1394
                |:  |          :   |   |  | :: :|::||:
Db       1793 LAKVLASIPDEHPLTAVVHAAGVLDDGVSESLTVERL 1829


RESULT 11
AAB70966
ID   AAB70966 standard; Protein; 2152 AA.
XX
AC   AAB70966;
XX
DT   28-AUG-2001  (first entry)
XX
DE   S. spinosa protein fragment encoded by ORF19, SEQ ID 44.
XX
KW   Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW   spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW   macrolide; insecticidal; polyketide synthase.
XX
OS   Saccharopolyspora spinosa.
XX
PN   DE19957268-A1.
XX
PD   08-MAR-2001.
XX
PF   29-NOV-1999;   99DE-1057268.
XX
PR   27-AUG-1999;   99DE-1040596.
XX
PA   (FARB ) BAYER AG.
XX
PI   Eberz G,  Moehrle V,  Froede R,  Velten R,  Salas JA;
XX
DR   WPI; 2001-267102/28.
DR   N-PSDB; AAF88336.
XX
PT   New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT   recombinant production of insecticidal spinosyns and their derivatives
PT   -
XX
PS   Claim 57; Page 200-209; 354pp; German.
XX
CC   This invention describes a novel method nucleic acid (I) and its encoded
CC   polypeptide (II) containing at least one region that encodes an enzymatic
CC   activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC   identify, inactivate or modulate genes involved in the biosynthesis of
CC   (II); (ii) to generate a library of polyketide synthases; (iii) for
CC   adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC   aglycone; and (iv) for recombinant production of the corresponding
CC   enzymes, which are used for production of (II), their precursors or
CC   derivatives, including production of transgenic plants that express (II)
CC   and thus have increased resistance to insects. (I) are also useful as
CC   markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC   macrolides with insecticidal, but not antibacterial, activity, and can
CC   also be used to raise specific antibodies, useful for identifying
CC   expression clones in a gene bank. Cells transformed with (I) may produce
CC   (II) at significantly increased levels or produce new derivatives of
CC   (II). This sequence represents an S. spinosa polyketide synthase.
XX
SQ   Sequence   2152 AA;

  Query Match             31.5%;  Score 2273;  DB 22;  Length 2152;
  Best Local Similarity   32.7%;  Pred. No. 5.3e-172;
  Matches  600;  Conservative  199;  Mismatches  542;  Indels  496;  Gaps   32;

Qy         11 EDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGKTP 69
              :||||||  ||| || |           | |  |  |  | :  || :| ||    | || | || :
Db         36 DDPIAIVAMSCRYPGQVSSPEDLWQLAAGGVDAISEVPGDRGWDLDGVFVPDSDRPGTSY 95

Qy         70 VTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETGV 129
                       || |: ||| ||||||||| |||  |||||| ||  | | :    :  |: ||
Db         96 ACAGGFLQGVSEFDAGFFGISPREALAMDPQQRLLLEVAWEVFERAGLEQRSTRGSRVGV 155

Qy        130 FIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSSL 189
              |:|     :| :  |       :    |     :| :||::|:  |    || | ||||  ||||
Db        156 FVGTNGQDYASWL--RTPPPAVAGHVLTGGAAAVLSGRVAYSFGFEGPAVTVDTACSSSL 213

Qy        190 VAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGRG 249
              ||:|||  |:||:|||   |||||::|  :|       |:    |:  |  |||||:|:|  ||| |  |
Db        214 VALHLAGQALRAGECDLALAGGVTVMSTPKVFLEFSRQRGLAPDGRCKSFAAGADGTGWG 273

Qy        250 EGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAGC 309
              ||   :::|:|||  || :|    :|||:||||:| ||||:||| ||||||:  |: :||| ||
Db        274 EGAGLLLLERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGSSQQRVITQALASAGL 333

Qy        310 AASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITG 369
               : |  |  |||||||| ||||||  ||| |  ||   ||      || :||||:|:||  : |:|: |
Db        334 SVSDVDAVEAHGTGTRLGDPIEAQALIATYGRDRDPGRPLWLGSVKSNIGHTQAAAGVAG 393

Qy        370 LLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSG 429
              ::|:|::::|||:|  ||  :: :|  : |          :  :         ||||      || ||||||:||
Db        394 VIKMVMAMRHGQLPRTLHVESPSPEVDWSAGTVQLLTENTPWPRSGRVRRVGVSSFGISG 453

Qy        430 TNAHVVLEEAPAATCTPPAP------ERPAELLVLSARTASALDAQAARLRDHLETYPSQ 483
              |||||:||: |               |       : |        ::| :|   || |||  | :|:  |
Db        454 TNAHVILEQPPGVPSQSAGPGSGSVVDVPVVPWMVSGKTPEALSAQATALMTYLDERPDV 513

Qy        484 CLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAF 543
                  || :|||  ||||.::  |    |   :  || |    :  | :  |    :  | |    :   |:: |
Db        514 SSLDVGYSLALTRSALDERAVVLGSDRETLLCGVKALSAGHEASGLVTGSVGAG-GRIGF 572

Qy        544 LFTGQGAQTLGMGRGLYDVWSAFREAFD-LCVRL---FNQELDRPLREVMWAEPASVDAA 599
              :|:|||  | |||||||||   :     |  |||   |    |        ||:     :|||:   .   :  ||
Db        573 VFSGQGGQWLGMGRGLYRAFPVFAAAFDEACAELDAHLGQEIG--VREVV----SGSDAQ 626

Qy        600 LLDQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVA 659
              |||:| :  |  ||   :    |    |  ||||  | :|  |||:|||  ||   |||  ||   |  |||
Db        627 LLDRTLWAQSGLFALQVGLLKLLDSWGVRPSVVLGHSVGELAAAFAAGVVSLSGAARLVA 686

Qy        660 ARGRLMQALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIA 719
               ||||||||:||  |:::: | |   :  :  :  : :|      ||||||      ||:::|     |
Db        687 GRARLMQALPSGGGMLAVPAGEELLWSLLADQGDRVGIAAVNAAGSVVLSGDRDVLDDLA 746

Qy        720 AAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKA-CTDE 778
                :   :| |:: | |||||||  | |||  |  :|  :| :|  |||    : :|| |:|
Db        747 GRLDGQGIRSRWLRVSHAFHSYRMDPMLAEFAELARTVDYRRCEVPIVSTLTGDLDDAGR 806

Qy        779 VSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDA----RPAL 834
              :| | ||||  || |||||||:||          | ||:||   |   |:  |:  :     | :
Db        807 MSGPDYWVRQVREPVRFADGVQALVEHDVATVVELGPDGALSALIQECVAASDHAGRLSA 866

Qy        835 LASSRAGRDEPATVLEALGGLWAVGGLVSWAGLFP-SGGRRVPLPTYPWQRERYW----- 888
              : :  |  :||   |': ||  :      ||| |  |     |  :| :::  |||| :||:|||
Db        867 VPAMRRNQDEAQKVMTALAHVHVRGGAVDWRSFFAGTGAKQIELPTYAFQRQRYWLVPSD 926

Qy        889 ------------------------------------------------------------ 888

Db        927 SGDVTGAGLAGAEHPLLGAVVPVAGGDEVLLTGRISVRTHPWLAEHRVLGEVIVAGTALL 986

Qy        889 --------------------------------------------------IDTKADDAAR 898
                                                                 : ::. : ||
Db        987 EIALHAGERLGCERVEELTLEAPLVLPERGAIQVQLRVGAPENSGRRPMALYSRPEGAAE 1046

Qy        899 GD-------RRAPGAGH------------------------------------------ 908
               |        | |||  |
Db       1047 HDWTRHATGRLAPGRGEAAGDLADWPAPGALPVDLDEFYRDLAELGLEYGPIFQGLKAAW 1106

Qy        909 ---DEV---------EEGG----------------------------------------- 915
                 |||      .    |: |
```

```
Db      1107 RQGDEVYAEAALPGTEDSGFGVHPALLDAALHATAVRDMDDARLPFQWEGVSLHAKAAPA 1166

Qy       916 ------------------------------------AVRGGDRRSA------RLDHP-- 930
                                                 | | | || |       || |
Db      1167 LRVRVVPAGDDAKSLLVCDGTGRPVISVDRLVLRSAAARRTGARRQAHQARLYRLSWPTV 1226

Qy       931 --------------------------------------------------PPESGRREKV 940
                                                                || ||  |
Db      1227 QLPTSAQPPSCVLLGTSEVSADIQVYPDLRSLTAALDAGAEPPGVVIAPTPPGGRTADV 1286

Qy       941 E----------------------------------------------------------- 941

Db      1287 RETTRHALDLVQGWLSDQRLNESRLLLVTQGAVAVEPGEPVTDLAQAALWGLLRSTQTEH 1346

Qy       942 ----------------------AAGD-----------------------------RPFR 949
                                   | |:                               :|
Db      1347 PDRFVLVDVPEPAQLLPALPGVLACGEPQLALRRGGAHAPRLAGLGSDDVLPVPDGTGWR 1406

Qy       950 LEIDEPGVLDHLVLRVTERRAP---GLGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPN 1006
             ||   || || | | | |  |     | ||| ||: |||::| |  :|||| |
Db      1407 LEATRPGSLDGLAL-VDEPTATAPLGDGEVRIAMRAAGVNFRDALIALGMYPG------- 1458

Qy      1007 PPLLLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEA 1066
                || | || :|  | || ||  |  |: :   ||            :|   |   |  |
Db      1459 -VASLGSEGAGVVVETGPGVTGLAPGDRVMGMIPKAFGPLAVADHRMVTRIPAGWSFARA 1517

Qy      1067 AAMPVAYLTAWYALDRIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKR 1126
             |::|: :|||:|||  :| |:|||  :|:|:|  ||||:||:|  |:|:||||:|||  : :
Db      1518 ASVPIVFLTAYYALVDLAGLRPGESLLVHSAAGGVGMAAIQLARHLGAEVYATASEDKWQ 1577

Qy      1127 AYLESLGVRYVSDSRSDRFVADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVEL 1186
             |   |    ::: ||:  |          |||  |||||||||:||    |   |:    |||:||
Db      1578 AV--ELSREHLASSRTCDFEQQFLGATGGRGVDVVLNSLAGEFADASLRMLPRGGRFLEL 1635

Qy      1187 GKRDCYADNQLG-LRPFLRNLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPI 1245
             || |       ::     | :    :|  |:              ||:  :| ||: |     |  | |:
Db      1636 GKTDVRDPVEVADAHPGVSYQAFDTVE------AGPQRIGEMLHELVELFEGRVLEPLPV 1689

Qy      1246 ATLPIARVADAFRSMAQAQHLGKLVLTLGDPEVQIRIPTHAGAGP-----STGDRDLLDR 1300
                 : :  :| | ::||:|:||||||:                |      ||         ||
Db      1690 TAWDVRQAPEALRHLSQARHVGKLVLTMP--------PVWDAAGTVLVTGGTG------- 1734

Qy      1301 LASAAPAARAAALEAFLRTQVSQVLRTPEIKVGAE--ALFTRLGMD-SLMAVELRNRIEA 1357
              |  |  ||    :|  :|   |    | |        ||  |   |  |  : || |  :: :| |
Db      1735 -ALGAEVARHLVIERGVRNLVLVSRRGPAASGAAELVAQLTAYGAEVSLQACDVADR-ET 1792

Qy      1358 SLKLKLSTTFLSTSPNIALLAQNLLDALATALSLERV 1394
               |:  |        :    |   | |  :: :|:||:
Db      1793 LAKVLASIPDEHPLTAVVHAAGVLDDGVSESLTVERL 1829

RESULT 12
AAE10142
ID   AAE10142 standard; Protein; 7068 AA.
XX
AC   AAE10142;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysI (partial).
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          34..448
FT                   /label= KS9_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          572..890
FT                   /label= AT9_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          904..1123
FT                   /label= DH9_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1443..1686
FT                   /label= KR9_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          1720..1793
FT                   /label= ACP9_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          1813..2236
FT                   /label= KS10_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2346..2664
FT                   /label= AT10_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2678..2890
FT                   /label= DH10_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          2983..3229
FT                   /label= KR10_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3266..3339
FT                   /label= ACP10_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3358..3780
FT                   /label= KS11_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          3898..4217
FT                   /label= AT11_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4231..4432
FT                   /label= DH11_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4523..4770
FT                   /label= KR11_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          4806..4879
FT                   /label= ACP11_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          4801..5325
FT                   /label= KS12_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          5432..5754
FT                   /label= AT12_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          5768..5977
FT                   /label= DH12_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          6068..6315
FT                   /label= KR12_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          6348..6421
FT                   /label= ACP12_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          6454..6873
FT                   /label= KS13_domain
FT                   /note= "Ketosynthase (KS) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
```

```
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17185.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 181-185; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei
CC   nystatin gene, NysI (partial).
XX
SQ   Sequence   7068 AA;

  Query Match            31.4%;  Score 2260.5;  DB 22;  Length 7068;
  Best Local Similarity  38.0%;  Pred. No. 3.5e-170;
  Matches  614;  Conservative  170;  Mismatches  544;  Indels  287;  Gaps   39;

Qy         10 AEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGKT 68
Db       4899 ADDPIAIVGMSCRYPGGVGSAEDLWRIALDEVDAISGFPADRGWDAEGLYDPDPDRPGRT 4958

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETG 128
Db       4959 YSVQGGFLRDVAEFDPGFFGISPREALSMDPQQRLLLETAWEAFEHAGIDPVGQRGSRTG 5018

Qy        129 VFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSS 188
Db       5019 TFVGASYQDYASGVPNSEGS---EGHMITGTLSSVLSGRVSYLFGFEGPAVTLDTACSSS 5075

Qy        189 LVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR 248
Db       5076 LVAMHLACQSLRNGESSLALAGGVSIMSTPMSFVGFSRQRALAEDGRCKAYADGADGMTL 5135

Qy        249 GEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAG 308
Db       5136 AEGVGLVLLERLSDARANGHQVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALANSA 5195

Qy        309 CAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
Db       5196 VAPGDIDVLEGHGTGTALGDPIEAQALLATYGQDRAPERPLLLGSVKSNIGHTQMASGVA 5255

Qy        369 GLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMS 428
Db       5256 SVIKLVRALQEGVVPKSLHIDRPSTHVDWSSGAIGLLTERTPWPETGRPRRAAVSSFGIS 5315

Qy        429 GTNAHVVLEEAPAATCTPPAPERPAELLV---LSARTASALDAQAARLRDHLETYPSQCL 485
Db       5316 GTNVHTILEQAPADEAPTPA-DPPRDGLVPVLLSGRGEAALRAQAARLLAFVEERPEAHL 5374

Qy        486 GDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLF 545
Db       5375 TDLAHSLATSRAALERRAAVIAADRDTLTRGLRALSDGRPDPGLVQGTA--GRGRTAFLF 5432

Qy        546 TGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTA 605
Db       5433 TGQGSQRPGMGRELHDRYPVFADALDEVLARLDDGPDRPLREVLFAAPDSAEAALLDRTG 5492

Qy        606 FTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLM 665
Db       5493 YAQPALFAVEVALFRLLTSWGLTPDYLAGHSVGELAAAHVAGVLSLDDACTLVAARGRLM 5552

Qy        666 QALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAAR 725
Db       5553 QALPEGGAMVALEAAEDEVLPLLEGLTDRVSVAAVNGPRSVVVAGVEEDVLLLADLFAAD 5612

Qy        726 GARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACT-DEVSSPGY 784
Db       5613 GRRTKRLRVSHAFHSPLMDAMLDDFAAVARGLTYHPPTIPFVSNVSGGLATAEQVRTPDY 5672

Qy        785 WVRHAREVVRFADGVKALHAAG-AGTFVEVGPKSTLLGLVPACMPD-ARPALLASSRAGR 842
Db       5673 WVGHVRAAVRFADGIDWLATQGDVHTFLELGPDGVLSAMARESLTDPSRTALLPTLRGDR 5732

Qy        843 DEPATVLEALGGLWAVGGLVSWAGLFPS-GGRRVPLPTYPWQRERYWIDTKA-------- 893
Db       5733 PEEPALVTAVAAAHAHGARVDWSGYFADHGARRTTLPTYAFQRERYWPDTTAATSAHTPG 5792

Qy        894 ------------------------DDAA-------------------------------- 897
Db       5793 SALDAEFWAAVERDDVAALAASLDLDDATVTAMVPALTAWRRRRGEQTELDSWRYRVTWK 5852

Qy        898 -RGDRRAPGA-----------GHDEVEEGGAVRGGDRRSA------RLDHPPPESG--R 936
Db       5853 PRGGATAPAALTGRWLVLVPHDHQDRQDDATAAWAADVETALGTTTVRLTVTTTDRAALA 5912

Qy        937 REKVEAAGDR-PFRLEIDEPGVLDHLVLRVTERRAPG----LGEVEIAVDAAGLSFNDVQ 991
Db       5913 ARITEAAGDQGPF------SGVLSLLPLATGDAGHPGAPAALTLTTTAVQALGDAGIDAP 5966

Qy        992 L--------ALG-----------------MVPDDLP----GKPNPPLLLGGECAGRIVA 1021
Db       5967 LWNVTRGAVAVGRAEQVTAPEQAAVWGLGRAVALELPARFGGTLDLPATLDGQAARRLRA 6026

Qy       1022 VGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALD 1081
Db       6027 VLAATD----GEDAVAL-----------------RPSGVFLRRLAHAPAGPDTARTAFD 6064

Qy       1082 RIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVH---------ATAGTPEKRAYLESL 1132
Db       6065 PAA-----GTVLITGGTGGIGGHVARRLARDGA-THLLLTSRRGPAAPGADALRAELEEL 6118

Qy       1133 GVR-------------------------------------YVSDSRS-DRFVADVRAW 1152
Db       6119 GARVTLAACDAADRDALAALLAELPDDAPLCAVFHTAGVVEDHVVDALTPENFAAVLRAK 6178

Qy       1153 TGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLRN-LSFSLV 1211
Db       6179 TVAAH---HLHELTADLDLAAFVLFSSTAGVLGAAGQGNYAAANAHLDALAEHRRSHGLT 6235

Qy       1212 DL--------RGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ 1262
Db       6236 ALSVAWGPWAGSGMVAD--------AAELTDRVRRGGF--EPLAPEP------AVRALLR 6279

Qy       1263 AQHLGKLVLTLGD---------------------PEVQIRIPTHAGAGPSTGDRDLLDR 1300
Db       6280 AIENDDTTVALADIDWERFQRAFAAVRPLPFVADLPETGRATPATA-TGAATG---LRQQ 6335

Qy       1301 LASAAPAARAAALEAFLRTQVSQVL-----RTPEIKVGAEALFTRLGMDSLMAVELRNRI 1355
Db       6336 LAELPEHERPAAVLDLLRTQVAAVLGHADPRTVE----DDHAFRDLGFDSLTILELRNAL 6391

Qy       1356 EASLKLKLSTTFLSTSPNIALLAQNLLDALA------TALSLERVAAENLRAGVQ 1404
Db       6392 NAATGLSLPATLVYDLPTPREMADFLLAELLGTLPTDTAATVASTASPKLSASFE 6446

RESULT 13
AAE10144
ID   AAE10144 standard; Protein; 9477 AA.
XX
AC   AAE10144;
XX
DT   29-NOV-2001  (first entry)
```

```
XX
DE   Streptomyces noursei nystatin gene, NysI (complete).
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          41..464
FT                   /label= KS15_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          578..889
FT                   /label= AT15_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          903..1102
FT                   /label= DH15_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1446..1731
FT                   /label= ER15_domain
FT                   /note= "Enoylreductase (ER) domain"
FT   Domain          1740..1988
FT                   /label= KR15_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          2023..2096
FT                   /label= ACP15_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          2117..2538
FT                   /label= KS16_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2635..2953
FT                   /label= AT16_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2967..3167
FT                   /label= DH16_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          3257..3500
FT                   /label= KR16_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3539..3612
FT                   /label= ACP16_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3634..4057
FT                   /label= KS17_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          4153..4472
FT                   /label= AT17_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4486..4725
FT                   /label= DH17_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4997..5245
FT                   /label= KR17_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          5277..5350
FT                   /label= ACP17_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   N-PSDB; AAD17186.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 15; Page 255-260; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a PKS type I encoding Streptomyces noursei
CC   nystatin gene, NysI (complete).
XX
SQ   Sequence   9477 AA;

  Query Match             31.4%;  Score 2260.5;  DB 22;  Length 9477;
  Best Local Similarity   38.0%;  Pred. No. 5.5e-170;
  Matches  614;  Conservative  170;  Mismatches  544;  Indels  287;  Gaps   39;

Qy         10 AEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAER-WDAAAWFDPDPDAPGKT 68
              |:||||||| ||| ||||        | :     | :   ||:| |||   :||||| ||:|
Db       4899 ADDPIAIVGMSCRYPGGVGSAEDLWRIALDEVDAISGFPADRGWDAEGLYDPDPDRPGRT 4958

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETG 128
                :  || ||| ||  ||||||||| |||  |||||  ||| |:| | |      |: ||
Db       4959 YSVQGGFLRDVAEFDPGFFGISPREALSMDPQQRLLLETAWEAFEHAGIDPVGQRGSRTG 5018

Qy        129 VFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSS 188
              |:|     :| : :| :   |     : |     ||: || :||:||  |   || | :||| |||
Db       5019 TFVGASYQDYASGVPNSEGS---EGHMITGTLSSVLSGRVSYLFGFEGPAVTLDTACSSS 5075

Qy        189 LVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR 248
              |||:|||||||||:|| | ||||||:| :| :   |: |||| ||||||::  |||
Db       5076 LVAMHLACQSLRNGESSLALAGGVSIMSTPMSFVGFSRQRALAEDGRCKAYADGADGMTL 5135

Qy        249 GEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAG 308
               ||   :|:|:||| |||:| ::||||||||:| |||| ||| ||  ||  |:::|||::
Db       5136 AEGVGLVLLERLSDARANGHQVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALANSA 5195

Qy        309 CAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
                |    :  :| ||||| |||||| ||| | || |       |||:||||:|:|| : |||:
Db       5196 VAPGDIDVLEGHGTGTALGDPIEAQALLATYGQDRAPERPLLLGSVKSNIGHTQMASGVA 5255

Qy        369 GLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMS 428
               ::|:| :|| | :| ||     :  : |      : :    |||||:   |||| ||||:|
Db       5256 SVIKLVRALQEGVVPKSLHIDRPSTHVDWSSGAIGLLTERTPWPETGRPRRAAVSSFGIS 5315

Qy        429 GTNAHVVLEEAPAATCTPPAPERPAELLV---LSARTASALDAQAARLRDHLETYPSQCL 485
              ||| |  :||:|||       || | : | : ||    || |   :|| ||||||    :|  |   |
Db       5316 GTNVHTILEQAPADEAPTPA-DPPRDGLVPVLLSGRGEAALRAQAARLLAFVEERPEAHL 5374

Qy        486 GDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLF 545
               |:| ||||:|:|:| ||| || |     |  | :  |:  || |:  |     ||: ||||
Db       5375 TDLAHSLATSRAALERRAAVIAADRDTLTRGLRALSDGRPDPGLVQGTA--GRGRTAFLF 5432

Qy        546 TGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTA 605
              ||||:|  |||| |:| :  | :| |  :  :  |||||||::| | | :|||||:|
Db       5433 TGQGSQRPGMGRELHDRYPVFADALDEVLARLDDGPDRPLREVLFAAPDSAEAALLDRTG 5492

Qy        606 FTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLM 665
               : ||||  | ||  |  |||: |: :||||:||| || |||| ||:||  ||||||||
Db       5493 YAQPALFAVEVALFRLLTSWGLTPDYLAGHSVGELAAAHVAGVLSLDDACTLVAARGRLM 5552
```

```
Qy       666 QALPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAAR 725
Db      5553 QALPEGGAMVALEAAEDEVLPLLEGLTDRVSVAAVNGPRSVVVAGVEEDVLLLADLFAAD 5612

Qy       726 GARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACT-DEVSSPGY 784
Db      5613 GRRTKRLRVSHAFHSPLMDAMLDDFAAVARGLTYHPPTIPFVSNVSGGLATAEQVRTPDY 5672

Qy       785 WVRHAREVVRFADGVKALHAAG-AGTFVEVGPKSTLLGLVPACMPD-ARPALLASSRAGR 842
Db      5673 WVGHVRAAVRFADGIDWLATQGDVHTFLELGPDGVLSAMARESLTDPSRTALLPTLRGDR 5732

Qy       843 DEPATVLEALGGLWAVGGLVSWAGLFPS-GGRRVPLPTYPWQRERYWIDTKA-------- 893
Db      5733 PEEPALVTAVAAAHAHGARVDWSGYFADHGARRTTLPTYAFQRERYWPDTTAATSAHTPG 5792

Qy       894 ------------------------DDAA------------------------------ 897
Db      5793 SALDAEFWAAVERDDVAALAASLDLDDATVTAMVPALTAWRRRRGEQTELDSWRYRVTWK 5852

Qy       898 -RGDRRAPGA------------GHDEVEEGGAVRGGDRRSA------RLDHPPPESG--R 936
Db      5853 PRGGATAPAALTGRWLVLVPHDHQDRQDDATAAWAADVETALGTTTVRLTVTTTDRAALA 5912

Qy       937 REKVEAAGDR-PFRLEIDEPGVLDHLVLRVTERRAPG----LGEVEIAVDAAGLSFNDVQ 991
Db      5913 ARITEAAGDQGPF------SGVLSLLPLATGDAGHPGAPAALTLTTAVQALGDAGIDAP 5966

Qy       992 L--------ALG------------------MVPDDLP----GKPNPPLLLGGECAGRIVA 1021
Db      5967 LWNVTRGAVAVGRAEQVTAPEQAAVWGLGRAVALELPARFGGTLDLPATLDGQAARRLRA 6026

Qy      1022 VGEGVNGLVVGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALD 1081
Db      6027 VLAATD----GEDAVAL------------------RPSGVFLRRLAHAPAGPDTARTAFD 6064

Qy      1082 RIARLQPGERVLIHAATGGVGLAAVQWAQHVGAEVH---------ATAGTPEKRAYLESL 1132
Db      6065 PAA-----GTVLITGGTGGIGGHVARRLARDGA-THLLLTSRRGPAAPGADALRAELEEL 6118

Qy      1133 GVR-----------------------------------------YVSDSRS-DRFVADVRAW 1152
Db      6119 GARVTLAACDAADRDALAALLAELPDDAPLCAVFHTAGVVEDHVVDALTPENFAAVLRAK 6178

Qy      1153 TGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLRN-LSFSLV 1211
Db      6179 TVAAH---HLHELTADLDLAAFVLFSSTAGVLGAAGQGNYAAANAHLDALAEHRRSHGLT 6235

Qy      1212 DL---------RGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ 1262
Db      6236 ALSVAWGPWAGSGMVAD--------AAELTDRVRRGGF--EPLAPEP------AVRALLR 6279

Qy      1263 AQHLGKLVLTLGD----------------------PEVQIRIPTHAGAGPSTGDRDLLDR 1300
Db      6280 AIENDDTTVALADIDWERFQRAFAAVRPLPFVADLPETGRATPATA-TGAATG---LRQQ 6335

Qy      1301 LASAAPAARAAALEAFLRTQVSQVL-----RTPEIKVGAEALFTRLGMDSLMAVELRNRI 1355
Db      6336 LAELPEHERPAAVLDLLRTQVAAVLGHADPRTVE----DDHAFRDLGFDSLTILELRNAL 6391

Qy      1356 EASLKLKLSTTFLSTSPNIALLAQNLLDALA------TALSLERVAAENLRAGVQ 1404
Db      6392 NAATGLSLPATLVYDLPTPREMADFLLAELLGTLPTDTAATVASTASPKLSASFE 6446
```

```
RESULT 14
AAU10701
ID   AAU10701 standard; Protein; 6095 AA.
XX
AC   AAU10701;
XX
DT   12-MAR-2002  (first entry)
XX
DE   Amino acid sequence encoded by S. cellulosum PKS cosmid pKOS28-26, ORF2.
XX
KW   Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW   acyl transferase domain; dehydratase domain; ketoreductase domain;
KW   acyl carrier protein domain; pKOS28-26.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   Domain          34..457
FT                   /label= Ketosynthase_domain
FT   Domain          568..895
FT                   /label= Acyltransferase_domain
FT   Domain          922..1100
FT                   /label= Dehydratase_domain
FT   Domain          1429..1614
FT                   /label= Ketoreductase_domain
FT   Domain          1706..1789
FT                   /label= Acyl_carrier_protein_domain
FT   Domain          1812..2235
FT                   /label= Ketosynthase_domain
FT   Domain          2341..2673
FT                   /label= Acyltransferase_domain
FT   Domain          2978..3162
FT                   /label= Ketoreductase_domain
FT   Domain          3250..3333
FT                   /label= Acyl_carrier_protein_domain
FT   Domain          3356..3781
FT                   /label= Ketosynthase_domain
FT   Domain          3887..4219
FT                   /label= Acyltransferase_domain
FT   Domain          4246..4424
FT                   /label= Dehydratase_domain
FT   Domain          4754..4939
FT                   /label= Ketoreductase_domain
FT   Domain          5031..5114
FT                   /label= Acyl_carrier_protein_domain
FT   Domain          5134..5558
FT                   /label= Ketosynthase_domain
FT   Domain          5664..5991
FT                   /label= Acyltransferase_domain
FT   Domain          6018..6095
FT                   /label= Dehydratase_domain
XX
PN   US6280999-B1.
XX
PD   28-AUG-2001.
XX
PF   31-AUG-1998;   98US-0144085.
XX
PR   22-JAN-1998;   98US-0010809.
XX
PA   (KOSA-) KOSAN BIOSCIENCE.
XX
PI   Gustafsson C,  Betlach MC,  Ashley G,  Julien B,  Ziermann R;
XX
DR   WPI; 2001-606536/69.
XX
PT   Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT   polyketide open reading frame encoding modules with one or more domains
PT   such as ketosynthase, acyl transferase and acyl carrier protein domains
PT    -
XX
PS   Claim 1; Fig 4; 72pp; English.
XX
CC   The present invention relates to the isolation of novel Sorangium
CC   cellulosum polyketide synthases (PKS), and the polynucleotide sequences
CC   encoding them. The polyketide synthases include catalytic domains such
CC   as ketosynthase domain, acyl transferase domain, dehydratase domain,
CC   ketoreductase domain and acyl carrier protein domain. A host cell
CC   comprising a PKS ORF (open reading frame) which encodes one or more
CC   more PKS domains is useful for producing polyketide synthases from which
```

```
CC   polyketides can be produced. The host cells are useful for constructing
CC   a library, where each individual colony of the library represents a
CC   colony with the ability to produce a particular PKS synthase and
CC   ultimately a particular polyketide. The polyketides produced by these
CC   colonies can be used collectively in a panel to represent a library or
CC   may be assessed individually for activity. Colonies in the library are
CC   also induced to produce the relevant synthases and thus to produce the
CC   relevant polyketides to obtain a library of candidate polyketides which
CC   can be screened for binding to desired targets such as receptors,
CC   signalling proteins, etc. The present sequence represents the amino acid
CC   sequence of encoded by ORF2 of S. cellulosum PKS cosmid pKOS28-26.
CC   Note: The present sequence is said to correspond to S. cellulosum PKS
CC   functional domains or domain subsets encoded by ORF2 of the cosmid
CC   pKOS28-26 (AAS17367).
XX
SQ   Sequence   6095 AA;

  Query Match             31.2%;  Score 2250;  DB 22;  Length 6095;
  Best Local Similarity   35.0%;  Pred. No. 1.9e-169;
  Matches  630;  Conservative  205;  Mismatches  550;  Indels  414;  Gaps   50;

Qy          7 ERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPG 66
Db       3351 EEGWSTPIAIVGVGLRMPGGASDLESFWQVLVEERDTLRPIPAQRFDVEALYDPDPDAKG 3410

Qy         67 KTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTE 126
Db       3411 KTYVRNASLLDDVASFDPGFFGISPREAEPMDPQHRLLLETAWSALEDAGVRPEHLKGSD 3470

Qy        127 TGVFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYS 186
Db       3471 TGVFVGVAPSEYAS---YRGKSANEDAYALTGTALSFAAGRVAYHLGLQGPAVSTDTACS 3527

Qy        187 SSLVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGF 246
Db       3528 SSLVAVHLACDALRRGDCEVALAAGVQVLANPAGFVLLSRTRALSPDGRCKAFSQAADGY 3587

Qy        247 GRGEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALAD 306
Db       3588 GRGEGVGVLVLMRLSEAQQQGKRVLGVVRGTAVNQDGASSGITAPNGTAQQKVVRAALRN 3647

Qy        307 AGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASG 366
Db       3648 AGLEPASIDVVECHGTGTSLGDPIEVQALGAVYGQGRDMARPLQLGAVKSNIGHLESAAG 3707

Qy        367 ITGLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWP--DWNTPRRAGVSS 424
Db       3708 IAGVCKILAAFRYESLPATLHSSPRNPRIPWENLPVQVVDRLTPWPRRAEGPPRRAGVSS 3767

Qy        425 FGMSGTNAHVVLEEAPAATCTPP--APERPAEL-LVLSARTASALDAQAARLRDHLETYP 481
Db       3768 FGISGTNAHVILEEAPAEARREPVEAEAAPALLPLVLSGRDEAAVNAQAGRWAKWLEEHG 3827

Qy        482 SQCLGDVAFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIADSSRGKL 541
Db       3828 EVGWSDVVRTAALHRTHFESRASVLAASAAGAVEGLRALSSGRPDAAVVSGTAKRG-GKL 3886

Qy        542 AFLFTGQGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALL 601
Db       3887 AVLFTGQGSQRLGMGKRLYEVYPVFRAAFDEVCEALDAHLDRGLREVVFAAAGSEEGAQL 3946

Qy        602 DQTAFTQPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAAR 661
Db       3947 ERTEYTQPGLFALEVALYRQWESWGLKPAALLGHSIGELSAAHVAGVLSLADAAKLVCAR 4006

Qy        662 GRLMQALPAGGAMVSIEAPEADVAAAVAPHAAS--VSIAAVNAPDQVVIAGAGQPVHAIA 719
Db       4007 GRLMQGCEAGGAMVSVEASEPEVQRALSEVGAQGRLSIAGLNAPMQTVLSGDEAAVLAVA 4066

Qy        720 AAMAARGARTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDE- 778
Db       4067 RRLEAQGRRTRRLRVSHAFHSAHMDGMLEEFGKVARECTYARPQLAVVSGVTGELGGEEA 4126

Qy        779 VSSPGYWVRHAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPD-ARPALLAS 837
Db       4127 LMSAEYWVRQVREAVRFLDGMRTLAAAGVSTYVECGPDGVLCALGAGCLPEGAEATFVAS 4186

Qy        838 SRAGRDEPATVLEALGGLWAVGGLVSWAGLFPS-GGRRVPLPTYPWQRERYWID------ 890
Db       4187 LRREQEEERALATAVATVHVQGHEVDWAQVLSGRGGRPVELPTYAFQRQRYWLEAPKART 4246

Qy        891 -----------------TK--------------------ADDAARGDRRAPGAGHDE- 910
Db       4247 DVGSAGLRESGHPLLGAATKLADGDGHLFTGRLSLGEQPWLRDHAVFGEVVFPGTGMLDL 4306

Qy        911 -VEEGGAVRGGDRRSARLDHP----------------PPESGRR------EKVEAAGDR 946
Db       4307 ALAAGRTVGSGALSELTISEPLMLAEDVAVRLQLSVGAPDAAGRRAFGLYSQPEQGPGDA 4366

Qy        947 PF-----RLEIDE----PGVLDHL--------------------------------VLR 964
Db       4367 PWVQHATGVLTDETLATSGELDELTTWPVPGAEAVDLSGFYERLHERGLRYGPAFQGLVE 4426

Qy        965 VTERRAPGLGEVEIAVDA---------------AGLSFNDVQLALGMVPDD--LP----- 1002
Db       4427 LSRRDATFFGRVVLPKDATDSAEDYGVHPALMDAALHTMVAAFAEVSAPDDVLLPFSWSD 4486

Qy       1003 ------GKPNPPL---LLGGECAGRIVAVGEGVNGLVVGQPVIALSAGAF---------- 1043
Db       4487 VALHATGASELRVRLELAGGRDSAQAAASLRVTD--AAGQPVV--SVGALHLRRATAEQL 4542

Qy       1044 --ATHV--------------------TTSAALVLPRPQALSAIEAAAMPVAYLTAWYALD 1081
Db       4543 RAATHAEAQHLYRVDFQLVSLVEAGSKVDSLVVLRAPEGRGRL-GEALGVEAIAGLDAL- 4600

Qy       1082 RIARLQPG----ERVLI----------------HAATG----------------GVGLA- 1104
Db       4601 -LARIEQGTRLPERVLVDMTAGSSQRSDMVISSHEATGQALSLLQAWLSEPRLEGVELVW 4659

Qy       1105 ----AVQWAQHVGAE--VHA-------TAGT--PEKRAYLESLGVRYVSD---------- 1139
Db       4660 VTRDAVSAAPGDGVQDLAHAPLWGLVRTARSEHPERRLRLIDVGTEPLDGGLLARALATA 4719

Qy       1140 --------------SRSDRFVADVRAWTGGEGVD----VVLNSLSGELIDK-SFNLLRSH 1180
Db       4720 TEPELALRGGAAMAARLVRVPAAAEGLTPARGLDPTGTVLVTGGTGELGQAVAEHLVRAH 4779

Qy       1181 G--RFVELGKRDCYADNQLGLRPFLRNLSFSLVDLRGMML---ERPARVRALLE------ 1229
Db       4780 GVRHLVLTSRRGLEAPGAPGFVQALEKLGAETVTVAACDVSKREEVARVLAGIEAAHPLT 4839

Qy       1230 ---ELLGLIAAGVF---TPPPIATLPIARV---------------------------- 1253
Db       4840 AVLHLAGVLDDGVITAQTPERLSRVLAPKVNGALHLHELTEDLDLSAFVLFSSMSGTLGT 4899

Qy       1254 ------------ADAFRS--------------------MAQAQHLGKLVLT-------- 1272
Db       4900 AGQSNYAAANSFLDAFAAHRRSRGLAATSLAWGFWAQTGVGMTAHLGEAELSRIQRAGLV 4959

Qy       1273 --------------LGDPEVQIRIPTH------------AGAGP--------------ST 1292
Db       4960 PIRVEEGLSLLDAALLRPEASL-VPAHLDLAQMQRGLEASGELPALLRALLRPGLRKASS 5018

Qy       1293 GDRD----LLDRLASAAPAARAAALEAFLRTQVSQVLRTPEIK-VGAEALFTRLGMDSLM 1347
Db       5019 ATRKEASALRERLSELPEAERLSSLVELVRAEVAAVLGLPRSEAVAVDQVLKDLGLDSLM 5078

Qy       1348 AVELRNRIEASLKLKLSTTFLSTSPNIALLAQNLLDALATALSLERVAAENLRAGVQND 1406
Db       5079 AVELRSRLSARAEIPLPATLVFDYPTPRAVAELL---LRQAFSKQQVTAARARRRTKED 5134

RESULT 15
AAY58577
ID   AAY58577 standard; Protein; 3798 AA.
XX
```

```
AC   AAY58577;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum type I polyketide synthase EPOS D.
XX
KW   EPOS D; type I polyketide synthase; epothilone biosynthesis;
KW   polyketide backbone formation; taxol substitute; anticancer.
XX
OS   Sorangium cellulosum.
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   N-PSDB; AAZ55887.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy  -
XX
PS   Claim 12; Page 139-150; 174pp; English.
XX
CC   This sequence represents a Sorangium cellulosum type I polyketide
CC   synthase, EPOS D, which is one of several epothilone biosynthetic
CC   enzymes encoded by a 68.75 kb contig. Epothilones A and B are
CC   16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence   3798 AA;

  Query Match            31.1%;  Score 2245.5;  DB 21;  Length 3798;
  Best Local Similarity  36.2%;  Pred. No. 2.1e-169;
  Matches  605;  Conservative  191;  Mismatches  495;  Indels  379;  Gaps   43;

Qy          9 AAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAAWFDPDPDAPGKT 68
Db       1519 AADDDIAIVGAACRFPGGDEGLETYWRHLAEGMVVSTEVPADRWRAADWYDPDPEVPGRT 1578

Qy         69 PVTRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCWEALENAAIAPSALVGTETG 128
Db       1579 YVAKGAFLRDVRSLDAAFFAISPREAMSLDPQQRLLLEVSWEAIERAGQDPMALRESATG 1638

Qy        129 VFIGIGPSEYEAALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCVAVDTAYSSS 188
Db       1639 VFVGMIGSEHAERVQGLDDDAAL-LYGTTGNLLSVAAGRLSFFLGLHGPTMTVDTACSSS 1697

Qy        189 LVAVHLACQSLRSGECSTALAGGVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR 248
Db       1698 LVALHLACQSLRLGECDQALAGGSSVLLSPRSFVAASRMRLLSPDGRCKTFSAAADGFAR 1757

Qy        249 GEGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNGSSQEIVLKRALADAG 308
Db       1758 AEGCAVVVLKRLRDAQRDRDPILAVVRSTAINHDGPSSGLTVPSGPAQQALLRQALAQAG 1817

Qy        309 CAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
Db       1818 VAPAEVDFVECHGTGTALGDPIEVQALGAVYGRGRPAERPLWLGAVKANLGHLEAAAGLA 1877

Qy        369 GLLKVVLSLQHGQIPAHLHAQALNPRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMS 428
Db       1878 GVLKVLLALEHEQIPAQPELDELNPHIPWAELPVAVVRRAVPWPRGARPRRAGVSAFGLS 1937

Qy        429 GTNAHVVLEEAPAATCTPPAPERPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDV 488
Db       1938 GTNAHVVLEEAPAVEPVAAAPERAAELFVLSAKSAAALDAQAARLRDHLEKHVELGLGDV 1997

Qy        489 AFSLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPGAVRSIAD-SSRGKLAFLFTG 547
Db       1998 AFSLATTRSAMEHRLAVAASSREALRGALSAAAQGHTPPGAVRGRASGGSAPKVVFVFPG 2057

Qy        548 QGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFT 607
Db       2058 QGSQWVGMGRKLMAEEPVFRAALEGCDRAIEAEAGWSLLGELSADEA---ASQLGRIDVV 2114

Qy        608 QPALFTFEYALAALWRSWGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQA 667
Db       2115 QPVLFAMEVALSALWRSWGVEPEAVVGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRR 2174

Qy        668 LPAGGAMVSIEAPEADVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGA 727
Db       2175 ISGQGEMALVELSLEEAEAALRGHEGRLSVAVSNSPRSTVLAGEPAALSEVLAALTAKGV 2234

Qy        728 RTKALHVSHAFHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVR 787
Db       2235 FWRQVKVDVASHSPQVDPLREELIAALGAIRPRAAAVPMRSTVTGGVIAGPELGASYWAD 2294

Qy        788 HAREVVRFADGVKALHAAGAGTFVEVGPKSTLLGLVPACMPDARP--ALLASSRAGRDEP 845
Db       2295 NLRQPVRFAAAAQALLEGGPALFIEMSPHPILVPPLDEIQTAAEQGGAAVGSLRRGQDER 2354

Qy        846 ATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPG 905
Db       2355 ATLLEALGTLWASGYPVSWARLFPAGGRRVPLPTYPWQHERYWI----EDSVHGSK---- 2406

Qy        906 AGHDEVEEGGAVRGGDRRSARLDHP----------------------------------- 930
Db       2407 ---------PSLRLRQLRNGATDHPLLGAPLLVSARPGAHLWEQALSDERLSYLSEHRVH 2457

Qy        931 ----PPESGRREKVEAAGDRPPFRLEIDEPG----VLDHLVL-----------RVTE---- 967
Db       2458 GEAVLPSAAYVEMALAAG-------VDLYGTATLVLEQLALERALAVPSEGGRIVQVALS 2510

Qy        968 RRAPGLGEVEIAV-DAAGLSFNDVQLALGMV---------------------PDDLPGK 1004
Db       2511 EEGPGRASFQVSSREEAGRSW--VRHATGHVCSGQSSAVGALKEAPWEIQRRCPSVLSSE 2568

Qy       1005 PNPPLL------------------LG-GECAGRIVAVGEGVNGLVVGQPVIALSAGAFAT 1045
Db       2569 ALYPLLNEHALDYGPCFQGVEQVWLGTGEVLGRVRLPGD-----------MASSSGAYRI 2617

Qy       1046 H-------------VTTSAALV-----------LPRP-----QALSAI---------- 1064
Db       2618 HPALLDACFQVLTALLTTPESIEIRRRLTDLHEPDLPRSRAPVNQAVSDTWLWDAALDGG 2677
```

```
Qy       1065 --EAAAMPV-----AYLTAWYALDRIAR------------LQPGERVLIHAATGGVGLA 1104
                ::|::||     ::    |  ::|:|:              |||  |       : ::
Db       2678 RRQSASVPVDLVLGSFHAKWEVMERLAQAYIIGTLRIWNVFCAAGERHTIDELLVRLQIS 2737

Qy       1105 AV------QWAQH-------VGAEVHATAG----TPEKRAYLESLGVRYVSDSRSDRFVA 1147
                |       :| :|       ||   |   :       |:  | || |         |  |
Db       2738 VVYRKVIKRWMEHLVAIGILVGDGEHFVSSQPLPEPDLAAVLEEAG----------RVFA 2787

Qy       1148 DVRA---WT--GGEGVDVVL-------------------------------NSLSGELI 1170
              |:      |     || :  ||                               | :   ::
Db       2788 DLPVLFEWCKFAGERLADVLTGKTLALEILFPGGSFDMAERIYRDSPIARYSNGIVRGVV 2847

Qy       1171 DKSFNLLRSHGRF--VELG---------------KRDCY-----------ADNQLGLR 1200
              : :  ::   | |  :|:|               |   |           |: :
Db       2848 ESAARVVAPSGMFSILEIGAGTGATTAAVLPVLLPDRTEYHFTDVSPLFLARAEQRFRDY 2907

Qy       1201 PFLRNLSFSLVD-----------------LRGMMLERPARVRALLEELLGLIAAGVF-- 1240
              |||:   : ::|                 :    ::      :||  :  || |:| |
Db       2908 PFLK---YGILDVDQEPAGQGYAHQRFDVIVAANVIHATRDIRATAKRLLSLLAPGGLLV 2964

Qy       1241 ----TPPPI----------------------ATLPIARVADAFRSMAQAQHL-----GK 1268
                  |  ||                        ||       |  |  :  |  :     |
Db       2965 LVEGTGHPIWFDITTGLIEGWQKYEDDLRIDHPLLPARTWCDVLRRVGFADAVSLPGDGS 3024

Qy       1269 LVLTLGDPEVQIRIPTHAGAG-PSTGDRDLLDRLASAAPAARAA--------------A 1312
                 ||    :  |  |  |||   |:|:        |: :|||||               |
Db       3025 PAGILGQHVILSRAPGIAGAACDSSGES------ATESPAARAVRQEWADGSADVVHRMA 3078

Qy       1313 LEA--FLRTQVSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLK 1360
              ||   | |     ||      :: |   | ||:       |:   |    :: |  ::
Db       3079 LERMYFHRRPGRQVWVHGRLRTGGGA-FTKALAGDLLLFEDTGQVVAEVQ 3127
```

Search completed: October 2, 2003, 17:33:51
Job time : 133.763 secs

OM protein - protein search, using sw model

```
Run on:         October  2, 2003, 17:27:29 ; Search time 37.1179 Seconds
                                             (without alignments)
                                             2912.986 Million cell updates/sec

Title:          US-10-014-717-8
Perfect score:  2124
Sequence:       1 MTQEQANQSETKPAFDFKPF..........HPAFRNIESLNVILKPSKAG 419

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:      830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_23:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2124	100.0	419	2	Q9L8C4	Q9l8c4 polyangium
2	2107	99.2	419	2	Q9KIZ4	Q9kiz4 polyangium
3	532.5	25.1	395	2	Q8KZM6	Q8kzm6 bacillus su
4	497	23.4	417	2	Q8RJZ4	Q8rjz4 stigmatella
5	488	23.0	416	2	O87605	O87605 streptomyce
6	486.5	22.9	415	16	Q92QQ5	Q92qq5 rhizobium m
7	486	22.9	376	16	O31785	O31785 bacillus su
8	483	22.7	399	2	Q93HA3	Q93ha3 streptomyce
9	481.5	22.7	418	16	Q9HYR4	Q9hyr4 pseudomonas
10	469	22.1	419	2	Q8KY32	Q8ky32 streptomyce
11	459	21.6	416	2	Q8KSY0	Q8ksy0 streptomyce
12	453.5	21.4	410	2	Q9X5P9	Q9x5p9 streptomyce
13	451	21.2	415	16	Q8UFY8	Q8ufy8 agrobacteri
14	447.5	21.1	401	2	Q9KHJ7	Q9khj7 streptomyce
15	446.5	21.0	394	2	Q8KNC8	Q8knc8 micromonosp
16	443	20.9	399	2	Q8GGQ1	Q8ggq1 streptomyce
17	442.5	20.8	434	16	Q9CBE7	Q9cbe7 mycobacteri
18	441	20.8	433	16	Q9F2Q0	Q9f2q0 streptomyce
19	433.5	20.4	399	2	Q93H81	Q93h81 streptomyce
20	431	20.3	379	2	Q93HF0	Q93hf0 streptomyce
21	430	20.2	411	2	Q60005	Q60005 streptomyce
22	430	20.2	417	2	Q59910	Q59910 streptomyce
23	429	20.2	411	16	Q9X8Q3	Q9x8q3 streptomyce
24	427.5	20.1	407	16	Q9RJQ7	Q9rjq7 streptomyce
25	427.5	20.1	416	2	Q93H80	Q93h80 streptomyce
26	425	20.0	511	2	O52569	O52569 amycolatops
27	423	19.9	397	2	Q8KIC8	Q8kic8 nocardia ae
28	421.5	19.8	444	16	Q9I107	Q9i107 pseudomonas
29	420	19.8	411	2	O32460	O32460 actinomadur
30	417	19.6	407	2	Q59819	Q59819 streptomyce
31	416.5	19.6	398	16	Q8XRX2	Q8xrx2 ralstonia s
32	413	19.4	397	2	Q59523	Q59523 micromonosp
33	412	19.4	397	16	Q8PHS9	Q8phs9 xanthomonas
34	411	19.4	404	2	Q54302	Q54302 streptomyce
35	409.5	19.3	416	16	Q982V9	Q982v9 rhizobium l
36	408	19.2	404	2	Q93HJ0	Q93hj0 streptomyce
37	407.5	19.2	436	2	Q9S4D6	Q9s4d6 streptomyce
38	406	19.1	310	2	O32927	O32927 mycobacteri
39	405.5	19.1	418	2	Q8GMF6	Q8gmf6 streptomyce
40	403.5	19.0	420	2	Q9ZHQ1	Q9zhq1 streptomyce
41	402.5	19.0	406	2	O87192	O87192 streptomyce
42	401.5	18.9	397	2	Q8KND6	Q8knd6 micromonosp
43	397.5	18.7	416	3	Q96WS9	Q96ws9 trichosporo
44	396.5	18.7	397	16	Q8P6C3	Q8p6c3 xanthomonas
45	393	18.5	399	2	Q93NX1	Q93nx1 streptomyce

ALIGNMENTS

```
RESULT 1
Q9L8C4
ID   Q9L8C4      PRELIMINARY;      PRT;   419 AA.
AC   Q9L8C4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Cytochrome P450 167A1.
GN   CYP167A1 OR EPOF.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SO CE90;
RX   MEDLINE=20130945; PubMed=10662695;
RA   Molnar I., Schupp T., Ono M., Zirkle R.E., Milnamow M.,
RA   Nowak-Thompson B., Engel N., Toupet C., Stratmann A., Cyr D.D.,
RA   Gorlach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RT   "The biosynthetic gene cluster for the microtubule-stabilizing agents
RT   epothilones A and B from Sorangium cellulosum So ce90.";
RL   Chem. Biol. 7:97-109(2000).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF210843; AAF26924.1; -.
DR   HSSP; P33006; 1CPT.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   419 AA;  46750 MW;  13E5857FB04FB8FE CRC64;

  Query Match             100.0%;  Score 2124;  DB 2;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-145;
  Matches  419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
```

```
Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300

Qy        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360

Qy        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 2
Q9KIZ4
ID   Q9KIZ4      PRELIMINARY;      PRT;   419 AA.
AC   Q9KIZ4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   EpoK.
GN   EPOK.
OS   Polyangium cellulosum.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Sorangineae; Polyangiaceae; Polyangium.
OX   NCBI_TaxID=56;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SMP44;
RX   MEDLINE=20293058; PubMed=10831849;
RA   Julien B., Shah S., Ziermann R., Goldman R., Katz L., Khosla C.;
RT   "Isolation and characterization of the epothilone biosynthetic gene
RT   cluster from Sorangium cellulosum.";
RL   Gene 249:153-160(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SMP44;
RX   MEDLINE=20115953; PubMed=10649995;
RA   Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C., Julien B.;
RT   "Cloning and heterologous expression of the epothilone gene cluster.";
RL   Science 287:640-642(2000).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF217189; AAF62886.1; -.
DR   HSSP; P33006; 1CPT.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   419 AA;  46750 MW;  FCCE0A193563688F CRC64;

  Query Match             99.2%;  Score 2107;  DB 2;  Length 419;
  Best Local Similarity   99.3%;  Pred. No. 1.4e-143;
  Matches  416;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60

Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240
              |||||||||:|||||||||||||||||||| |||||||||||||||||||||||||||||
Db        181 RALGVGLVPRVDEETKTLVASVTEGLALLHGVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300

Qy        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTSASLAYGR 360

Qy        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 3
Q8KZM6
ID   Q8KZM6      PRELIMINARY;      PRT;   395 AA.
AC   Q8KZM6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cytochrome P450 enzyme.
GN   BIOI.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OK2;
RA   Sasaki M., Kotanagi T., Kurusu Y.;
RT   "Genetic analysis of biotin operon in Bacillus subtilis natto OK2.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AB088066; BAC03244.1; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   395 AA;  44869 MW;  6D4B5C445107D225 CRC64;

  Query Match             25.1%;  Score 532.5;  DB 2;  Length 395;
  Best Local Similarity   32.1%;  Pred. No. 3.3e-30;
  Matches  130;  Conservative   96;  Mismatches  138;  Indels   41;  Gaps   11;

Qy         24 YAEDPFPAIERLREATPIFYWDEGR-----SWVLTRYHDVSAVFRDERFAVSREEWESSA 78
              : ::|:  :  ||   ||:.  :|       | :| | : :|: :| || |     ||| 
Db         12 FLKNPYSFYDTLRAVHPIY---KGSFLKYPGWYVTGYEETAAILKDARFKVRTPLPESST 68

Qy         79 EYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARS 138
              :|     :|| ::   :      || |:| | : :|| || :  :  |  || ||||   
Db         69 KYQ----DLSHVQNQMMLFQNQPDHRRLRTLASGAFTPRATESYQPYIIETVHQLLDQVQ 124

Qy        139 GQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTL 198
              |::: :|: |:|  :    |: :: || |  |: : : ::        |:  :|      
Db        125 GEKKMEVISDFAFPLASFVIANIIGVPEEDREQLKEWAAS--------LIQTIDFTRSRK 176

Qy        199 VASVTEG-------LALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAI 251
              |  :|||       :|   ::: :|:|:| : |:::|||: : :  :|: :|       :
Db        177 V--LTEGNHMAVQAMAYFKELIQKRKRHP-QQDMISMLLKGK-ENDKLTEEEAASTCILL 232

Qy        252 IAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQ 311
                || :||: ||: :|| ||: || |  ::  | |:  |:| ||:::  :: | | | :
Db        233 AIAGHETTVNLISNSVLCLLQHPEQLLKLRENPDLIGTAVEECLRYESPTQM-TARVASE 291
```

```
Qy        312 DLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLA 371
              |::  | :|::|| |:||: :| || ::|: |||||: |      |::| | ||| | |||
Db        292 DIDISGVTIRQGEQVYLLLGAANRDPSIFTNPDVFDITRSPNPHLSFGHGHHVCLGSSLA 351

Qy        372 RLEAEIAVGTIFRRFPEMKLKE-----TPVFGYHPAFRNIESLNV 411
              ||||:||: |  :| | :|| :      |:||    || :| | |
Db        352 RLEAQIAINTFLQRMPSLKLADFEWRYRPLFG----FRALEELPV 392


RESULT 4
Q8RJZ4
ID   Q8RJZ4      PRELIMINARY;      PRT;   417 AA.
AC   Q8RJZ4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cytochrome P450 monooxygenase.
OS   Stigmatella aurantiaca.
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC   Cystobacterineae; Cystobacteraceae; Stigmatella.
OX   NCBI_TaxID=41;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sg a15;
RA   Gaitatzis N., Silakowski B., Kunze B., Nordsiek G., Blocker H.,
RA   Hofle G., Muller R.;
RT   "The biosynthesis of the aromatic myxobacterial electron transport
RT   inhibitor stigmatellin is directed by a novel type of modular
RT   polyketide synthase.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AJ421825; CAD19077.1; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   417 AA;  45973 MW;  E8F74DFD7BF6A848 CRC64;

  Query Match             23.4%;  Score 497;  DB 2;  Length 417;
  Best Local Similarity   33.3%;  Pred. No. 1.3e-27;
  Matches  136;  Conservative   69;  Mismatches  188;  Indels   16;  Gaps    8;

Qy         14 AFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREE 73
              |||      :  ||||    :||  |:    :| |  |||: |:  ::|| |    | :
Db         19 AFDLARLDDAFYADPFPLYRAMRERDPVHRMPDG-SLFLTRWADLDCIYRDTRTFSSDKR 77

Qy         74 WESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQL 133
              | |:|  |  |                :|  ||  |||:|:  :  | ||:  :   :: |:|
Db         78 AEFGAKYGDA-PLFEHHTTSLVFNDPPL-HTRVRRLIVGALTPRALSTMEPGLRTLVDRL 135

Qy        134 LDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDE 193
              ||  : :   |:: |:|  ||:  |  || :| |     |  :  |    |   ||::
Db        136 LDGLAVKGAADLIEDFAAAIPIEVIGNLLDIPTEERGPLRGWSLAILGA----LEPRLTA 191

Qy        194 ETKTL-VASVTEGLALLHDVLDERRRNPLE--NDVLTMLLQAEADGSRLSTKELVALVGA 250
              | :     :||| |  |  :: :||  | :    |||| |:| |:|| ||: ||:
Db        192 EQEACGNEAVTEFLDYLRILVAQRRARPGDPATDVLTRLIQGESDGERLTETELLHQCVF 251

Qy        251 IIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFAR 310
              :: || :|| ||  |:  | | |:    :  | |:  |::||||::: ::| |   | |
Db        252 LLNAGHETTTNLIGNALELLARFPDERARLLRAPALIPTAVEEVLRYESSNQLGNRRVA- 310

Qy        311 QDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSL 370
              :| |  | ::  |  :  | | | || |  |  | |: ||||   ||:    | |  | ||::|
Db        311 EDTEIGGVAVPTGTFLTLCIGAANRDPARFEDPEHFDVGRQPNRHLAFAGGAHTCAGMNL 370

Qy        371 ARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              ||:|| ||:      |||:  |   ||     ||    ::       |:: |
Db        371 ARIEARIALAAFLARFPDYALTAPPVRARRARFRGFTAM-----PTRLG 414


RESULT 5
O87605
ID   O87605      PRELIMINARY;      PRT;   416 AA.
AC   O87605;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cytochrome P450 CYP107L1.
GN   CYP107L1 OR PICK OR PIKC.
OS   Streptomyces violaceus (Streptomyces venezuelae).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1936;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC15439;
RA   Betlach M.C., Kealey J.T., Betlach M.C., Ashley G.W., McDaniel R.;
RT   "Characterization of the macrolide P450 hydroxylase from Streptomyces
RT   venezuelae which converts narbomycin to picromycin.";
RL   Biochemistry 0:0-0(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC15439;
RX   MEDLINE=99051447; PubMed=9831532;
RA   Xue Y., Wilson D., Zhao L., Liu H.-W., Sherman D.H.;
RT   "Hydroxylation of macrolactones YC-17 and narbomycin is mediated by
RT   the pikC-encoded cytochrome P450 in Streptomyces venezuelae.";
RL   Chem. Biol. 5:661-669(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC15439;
RX   MEDLINE=98445333; PubMed=9770448;
RA   Xue Y., Zhao L., Liu H.W., Sherman D.H.;
RT   "A gene cluster for macrolide antibiotic biosynthesis in Streptomyces
RT   venezuelae: architecture of metabolic diversity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF087022; AAC64105.1; -.
DR   EMBL; AF079139; AAC68886.1; -.
DR   HSSP; Q00441; 1OXA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   416 AA;  46038 MW;  B7392C742045F06B CRC64;

  Query Match             23.0%;  Score 488;  DB 2;  Length 416;
  Best Local Similarity   30.5%;  Pred. No. 5.6e-27;
  Matches  128;  Conservative   82;  Mismatches  185;  Indels   24;  Gaps    8;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHD 57
              | :  |   :  : |  |        :| ||:|  |||  |         ||   ||: : |
Db          1 MRRTQQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDR 60

Qy         58 VSAVFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSR 117
              || | ||:   ::| :|   :: : |       : :      |  | |:||||  ||| |
Db         61 ARAVLADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMR 114

Qy        118 AIDLLRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRF 175
              ::|||  :|  || |:||      :    |::  |   :|: || || ||        || :
Db        115 RVELLRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVW 174

Qy        176 GSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEAD 235
              |              : |   : :| :| ::  |: |   :| :|    |: :  :: :: |
Db        175 TDAF-------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDED 224

Qy        236 GSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVL 295
              ||||:::||: :   :: || :||: |||  :   || |:  | ::|:  |:  ||:|:|
Db        225 GSRLTSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEML 284

Qy        296 RFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGAS 355
              |::  :   | || : ::  |   |  |: | ::: ||  |        |  | ||:||||
Db        285 RYEGPVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGH 344
```

```
Qy        356 LAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
              ||:| | | | |  |||||| |||. :  | |:: |  :|    |:  :|  | :::| :
Db        345 LAFGHGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403


RESULT 6
Q92QQ5
ID   Q92QQ5      PRELIMINARY;      PRT;   415 AA.
AC   Q92QQ5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative cytochrome P450-like monooxygenase protein
DE   (EC 1.14.14.1).
GN   R01252 OR SMC01812.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396507; PubMed=11481430;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR   EMBL; AL591786; CAC45831.1; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Complete proteome.
SQ   SEQUENCE   415 AA;  46750 MW;  8696BA61B71B14E7 CRC64;

  Query Match             22.9%;  Score 486.5;  DB 16;  Length 415;
  Best Local Similarity   31.8%;  Pred. No. 7.2e-27;
  Matches  127;  Conservative   77;  Mismatches  179;  Indels   17;  Gaps   10;

Qy         22 PGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERFA-----VSREEWES 76
              | :  :| ||     |      | |:|:| : |    |  |::: || ||       |:  |
Db         23 PRFFRNPLPAYAALHAQCPAFFWEEPQQWFFAGYEQVNSLLRDRRFGRQILHVATREELG 82

Qy         77 SAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDA 136
               |   : :    : : :  | |   | |:| ||| :| || |: || ||:        ::|
Db         83 MPEPKPHLKDFDALEAHSLLELEPPAHTRLRTLVNRAFVSRQIEELRPEIEALSHAVIDG 142

Qy        137 RSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETK 196
                  | ::::  ||| ||:  |: :| :| |    :     :       :      | :: |
Db        143 FEKDGETELLKTYAETIPVTIIARMLGIPVEAAPRLLDWSHRMVKMYVFN--PSLETEFD 200

Qy        197 TLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGT 256
                 ||  |     |  :: |:| || : |:|| :: :| || ||| ||::     :: ||
Db        201 ANNAS-AEFADYLKGIIAEKRTNPAD-DLLTHMITSEKDGERLSDAELISTTVLLLNAGH 258

Qy        257 DTTIYLIAFAVLNLLRSP-EALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEY 315
              : |:: |  ||  :|:|        ||       |     :: |  |||   ||   |:| |:|
Db        259 EATVHQIGNAVRTILQSGLSPAELFSDEKATER-TVEECLRFAAPLHI-FQRYALMDIEL 316

Qy        316 -CGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLE 374
                | :::||: :  |:: :|   |   || || |   |: ||:::::| | | | |  |||||
Db        317 ENGIALRKGDKIGLMLGAANVDPRKFSSPDTFRPDRNEGANVSFGAGLHFCIGAPLARLE 376

Qy        375 AEIAVGTIFRRFPEMKLKETPVF--GYHPAFRNIESLNVI 412
               :|::  :||| | :|| |      :|  |  :| |:::
Db        377 LQISLPILFRRLPGMRLKNEPPVKDAFH--FHGLERLDLV 414


RESULT 7
O31785
ID   O31785      PRELIMINARY;      PRT;   376 AA.
AC   O31785;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative cytochrome P450 107K1.
GN   PKSS OR CYP107K1.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; Z99113; CAB13607.1; -.
DR   HSSP; Q00441; 1OXA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Hypothetical protein; Monooxygenase; Oxidoreductase;
KW   Complete proteome.
SQ   SEQUENCE   376 AA;  43418 MW;  C3899600FFD87ECD CRC64;

  Query Match             22.9%;  Score 486;  DB 16;  Length 376;
  Best Local Similarity   31.6%;  Pred. No. 6.8e-27;
  Matches  120;  Conservative   76;  Mismatches  156;  Indels   28;  Gaps    7;

Qy         17 FKPFAPGYAEDPFPAIERLREATPIFYWDEGR------SWVLTRYHDVSAVFRDERFA-- 68
              | |     :  :||  : | | ||  ||  :: |         :|::||| |  | :| |
Db          8 FHPHGKEFHHNPFSVLGRFREEEPIHRFELKRFGATYPAWLITRYDDCMAFLKDNRITRD 67

Qy         69 ---VSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAE 125
                 |  :|        |  |  :||      :          || |:| ||: :|| |  |: ||
Db         68 VKNVMNQEQIKMLNVSEDIDFVSD----HMLAKDTPDHTRLRSLVHQAFTPRTIENLRGS 123
```

```
Qy        126 IQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARAL-- 183
              |::  :||||        : : |::: :|    :|    || |: :|  |     :|: : :|
Db        124 IEQIAEQLLDEMEKENKADIMKSFASPLPFIVISELMGIPKEDRSQFQIWTNAMVDTSEG 183

Qy        184 GVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKE 243
                |    |   |   |   :|         |:||        ||  | ::|:::  |: || :||:|| ||
Db        184 NRELTNQALREFKDYIAK------LIHD----RRIKP-KDDLISKLVHAEENGSKLSEKE 232

Qy        244 LVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRI 303
              | :::   :: || :||: |:       | ||:   :   | :| :| ::    |::|:||: : : :
Db        233 LYSMLFLLVVAGLETTVNLLGSGTLALLQHKKECEKLKQQPEMIATAVEELLRYTSPVVM 292

Qy        304 GTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPH 363
                  |:| :|   |  | |||:|:|:|:|: | || ||     |   |:: :: |       :::| |  |
Db        293 MANRWAIEDFTYKGHSIKRGDMIFIGIGSANRDPNFFENPEILNINRSPNRHISFGFGIH 352

Qy        364 VCPGVSLARLEAEIAVGTIF 383
                  | |   |||||   ||       |
Db        353 FCLGAPLARLEGHIAFKAAF 372
```

RESULT 8
Q93HA3

```
ID   Q93HA3      PRELIMINARY;      PRT;   399 AA.
AC   Q93HA3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cytochrome P450.
OS   Streptomyces avermitilis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AB070947; BAB69286.1; -.
DR   HSSP; P23295; 1JFB.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   399 AA;  44538 MW;  CB0D90378008F99D CRC64;

  Query Match             22.7%;  Score 483;  DB 2;  Length 399;
  Best Local Similarity   32.2%;  Pred. No. 1.2e-26;
  Matches  126;  Conservative   66;  Mismatches  163;  Indels   36;  Gaps    7;

Qy         24 YAEDPFPAIERLREATPIFYWDEGR---SWVLTRYHDVSAVFRDERFAVSRE----EWES 76
               :  ||:|      ||     |:      |||        ::|||  ||   :   |   |  :              :|:
Db          3 FGADPYPEYAWLRAEEPVRQVLEGRGLYGLLVTRYEDVRKLLSDPRMSKDPRNAPLDWQE 62

Qy         77 SAEYSSAIPELSDMKKYG--LFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLL 134
              : :               |  |     |  |           :|  |:|:||: :||:|  :: |||::|     |  ||
Db         63 AGKGR----PLEDRTGLGTHLLTTDAPEHTRLRRLVSTAFTARRVEGLRAQVQHITDGLL 118

Qy        135 DARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEE 194
              |      : : :::  |:|   : :    |   || ||      : ||::            |
Db        119 DTIVPRGQAELIGDFAFPLAITVICELLGVPKADQDVFRQWTKDFRR-----------W 166

Qy        195 TKTLVASVTEG---------LALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKEL 244
              |  |    |                  |   ::|:|:|::|  :   ::    |:  |     |    ||:   ||
Db        167 TNTDSAQADRGDARPVGLRDLLEYLTRLVDKRRQDPADG-LVDALIAARDDDDRLNEAEL 225

Qy        245 VALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIG 304
```

```
                 ::::   ::   | :||: ||       | |||  |:  |  |::    |  |:  :||:|:||:|
Db        226 LSMMSLLLVGGFETTVNLIGNGTLALLRHPDQLALLRERPELVDSALEEMLRYDGSFETA 285

Qy        305 TVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHV 364
               | ||  : :|    |  |:||   | |   :  || |||     | || |||  |       | :|:||| |
Db        286 TWRFPLEPIEVAGTRIEKGHPVLLSLASANRDGAKFPAPDDFDVTRADPAHVAFGRGAHF 345

Qy        365 CPGVSLARLEAEIAVGTIFRRFPEMKLKETP 395
              |  |    |||||   ||       :  || |  :  |        |
Db        346 CLGAPLARLEGRIAFHGLLRRLPGLALSVPP 376
```

RESULT 9
Q9HYR4

```
ID   Q9HYR4      PRELIMINARY;      PRT;   418 AA.
AC   Q9HYR4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative cytochrome P450 107S1.
GN   CYP107S1 OR PA3331.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AE004755; AAG06719.1; -.
DR   HSSP; Q00441; 1OXA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
KW   Heme; Hypothetical protein; Monooxygenase; Oxidoreductase;
KW   Complete proteome.
SQ   SEQUENCE   418 AA;  46377 MW;  8231E591DB10F351 CRC64;

  Query Match             22.7%;  Score 481.5;  DB 16;  Length 418;
  Best Local Similarity   33.5%;  Pred. No. 1.7e-26;
  Matches  116;  Conservative   73;  Mismatches  140;  Indels   17;  Gaps    5;

Qy         50 WVLTRYHDVSAVFRDERFAVSREEWESSAEYS----SAIPELSDMKKYGLFGLPPEDHAR 105
              ||:|||  |       |      | |:  :::   |:      |          :  :     :  :   | | || |
Db         41 WVVTRYRDARKVLNHP--GVRRDARQAAELYAKRTGSPRAGIGEGLSHHMLNLDPPDHTR 98

Qy        106 VRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVP 165
              :| ||   :|| |  :: |:   |:|   :  ||||  :|:|:  |:: |:|   :   :    :      || :|
Db         99 LRSLVGRAFTPRQVERLQPHIERITEALLDAMAGREQADLMADFAIPLTIAVIFELLGIP 158

Qy        166 AECDEKFRRFGSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDV 225
                         | |:              | |     |      ||  :    :  :   :|    |: :||  | :| ||
Db        159 EAEREHARQSWERQAELLS-------PEEAQALADAQVDYLRVL---LEAKRRQPAD-DV 207

Qy        226 LTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPG 285
               : |:||   :  :||  |||::     :: :|  :||: :|    |:: || :|| |  |::|:|
Db        208 YSGLVQAADESGQLSEAELVSMAHLLMMSGFETTMNMIGNALVTLLVNPEQLALLRAQPE 267

Qy        286 LMRNALDEVLRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDV 345
              |: ||::|::| |: :|    :||   :|:|    |  :|    || :  :     :|  |       |    ||
Db        268 LLPNAMEELVRHDSPVRASMLRFTVEDVELDGVTIPAGEYILVSNLTANHDAERFDDPDR 327

Qy        346 FDVRRDTGASLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKL 391
               |: |:|    | ||  |  |  |  |  ||||||   ||:   :    |||:::|
```

```
Db        328 LDLTRNTDGHLGYGFGVHYCVGASLARLEGRIAIQRLLARFPDLQL 373

RESULT 10
Q8KY32
ID   Q8KY32      PRELIMINARY;      PRT;   419 AA.
AC   Q8KY32;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   RubU.
GN   RUBU.
OS   Streptomyces collinus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=42684;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM2012;
RA   Saito H., Bruenker P., Martin R., Minas W.;
RT   "Streptomyces collinus DSM2012 rubromycin biosynthesis gene cluster.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF293355; AAM97370.1; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   419 AA;  46331 MW;  361266AAFE79DE5B CRC64;

  Query Match             22.1%;  Score 469;  DB 2;  Length 419;
  Best Local Similarity   30.6%;  Pred. No. 1.3e-25;
  Matches  121;  Conservative   72;  Mismatches  169;  Indels   34;  Gaps    6;

Qy         20 FAPGYAEDPFPAIERLRE---ATPIFYWDEGRSWVLTRYHDVSAVFRDERFA----VSRE 72
              | | : ::|  |:  ||     | |:   :  |:|::| :     |:  | | :     | |:
Db         16 FTPAFHQNPHEALAGLRRTAPAVPVMTPNGLRTWLVTGHEHARALLADPRLSKDMRVGRD 75

Qy         73 -------------EWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAI 119
                           |: : :    |   |  :   :     :     | || |:|:||  :||:| :
Db         76 LIPRNFVDPDKQREFLAESGERSQFPHVLSVH---MLDSDPPDHTRLRRLVGRAFTARRV 132

Qy        120 DLLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179
              : ||  |     |:||||  :   |  |::    |   :|     |  || ||  :      |||: :
Db        133 ESLRPRITELTDELLDAMARHERLDLMEALAFPVPFTVICWLLGVPPDDRAAFRRWSNLL 192

Qy        180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRL 239
                    |      :|  | : ::: :||          |  | :.||    :|:||  |:   |    :|
Db        193 VSGAG---TDEVREASASMITYLTE--------LIEAKRNEPADDMLTDLVHARDAGDQL 241

Qy        240 STKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDN 299
              |: ||:::    :: || :||: ||      | ||   ||    |  : |:     |::|  ||:|
Db        242 SSDELISMAFLLLVAGHETTVNLIGNGALALLTHPEVREQLAADESLWPGAVEEFLRYDG 301

Qy        300 ILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYG 359
                 :      | ||  : :|       :| :|| | : | :| ||     :    ||   |: |        |:|:|
Db        302 PVTNATWRFTTEPVEVGSVTIPEGEFVTISIGAAGRDPDRYPDPDRLDITRAHSGSVAFG 361

Qy        360 RGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP 395
                 | | | |   |||||  |  :     :| |  |  :: |      |
Db        362 HGIHHCLGAPLARLEGRIVLSRLFARLPGLRLAADP 397

RESULT 11
Q8KSY0
ID   Q8KSY0      PRELIMINARY;      PRT;   416 AA.
AC   Q8KSY0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative cytochrome P450 GrhO3.
GN   GRHO3.
OS   Streptomyces sp. JP95.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=195040;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JP95;
RA   Li A., Piel J.;
RT   "A gene cluster from a marine Streptomyces encoding the biosynthesis
RT   of the aromatic spiroketal polyketide griseorhodin A.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF509565; AAM33670.1; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   416 AA;  46093 MW;  4E33B40A11EBD4D9 CRC64;

  Query Match             21.6%;  Score 459;  DB 2;  Length 416;
  Best Local Similarity   31.8%;  Pred. No. 6.9e-25;
  Matches  134;  Conservative   76;  Mismatches  154;  Indels   58;  Gaps   13;

Qy          3 QEQANQSETKPAFDFK---PF--APGYAEDPFPAIERLREATPI---FYWDEGRSWVLT- 53
              |  |: :|  ||:  |     |:     |||        | |||  ||       |:     :|::|
Db          9 QTLADGAEDAPAYPVKRTCPYRMPPGY--------EELREKGPISRVTLWNGRTAWLVTG 60

Qy         54 ------RYHD--VSAVFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHAR 105
                     : |  :|:   | || :      |:  : ::| |               | |:     |||
Db         61 NDLGRRLFPDARLSSDVLDPRFPLLAPRIEAQRQQAAAPP---------LVGVDDPVHAR 111

Qy        106 VRKLVNPSFTSRAIDLLRAEIQRTVDQLLD---ARSGQEEFDVVRDYAEGIPMRAISALL 162
                |::| |||  | |: || |||:  | |||    |:          |:: :||  :|    |  ||
Db        112 QRRMVLPSFGIRQINALRPEIQKYADDLLDTMLAKGPGVTVDLLTEYALPMPSAVICMLL 171

Qy        163 KVPAE----CDEKFRRFGSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRR 218
                || |       ||: |    |::              :|:        : || || | |::  ::
Db        172 GVPYEDHHYFDERSRHVLSSSG-----------EEQAAQAQQAFTEILAYLDDLIVRKQA 220

Qy        219 NPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALE 278
                |  : :|    :  :  :   ::  :||   :     :: :| :||  :||  : : || |: |
Db        221 EP-GDTLLDELIARQLEEGKVDRQELAMIATVLLVSGHETTSNMIALSTMALLADPDQLA 279

Qy        279 LVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGT 338
                ::|:  ||  |:||::|| :|    :  :| |::|:|   |    |: |: |  |        ||
Db        280 ALRADESLMPRAVDELMRFSSIGDM-LMRVAKEDIEIEGHLIRAGDGVILSTMLMNRDPG 338

Qy        339 VFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKL----KET 394
                | |||  |:||   |    :|:| | | | | :||| | |||: |:||| | :||        ::
Db        339 AFERPDELDIRRPAGRHVAFGYGIHQCIGQNLARAEMEIALATLFRRVPTLKLAVPAEQV 398

Qy        395 PV 396
                ||
Db        399 PV 400

RESULT 12
Q9X5P9
ID   Q9X5P9      PRELIMINARY;      PRT;   410 AA.
AC   Q9X5P9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative cytochrome P450 107N1.
GN   CYP107N1.
OS   Streptomyces lavendulae.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1914;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2564;
RX   MEDLINE=99201491; PubMed=10099135;
```

```
RA   Mao Y.Q., Varoglu M., Sherman D.H.;
RT   "Molecular characterization and analysis of the biosynthetic gene
RT   cluster for the antitumor antibiotic mitomycin C from Streptomyces
RT   lavendulae NRRL 2564.";
RL   Chem. Biol. 6:251-263(1999).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF127374; AAD28449.1; -.
DR   HSSP; Q55080; 1IO7.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Hypothetical protein; Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   410 AA;  45887 MW;  1B9C78D8AD34E748 CRC64;

  Query Match             21.4%;  Score 453.5;  DB 2;  Length 410;
  Best Local Similarity   32.5%;  Pred. No. 1.7e-24;
  Matches  134;  Conservative   61;  Mismatches  140;  Indels   77;  Gaps   13;

Qy         19 PF----APGYAEDPFPAIERLREATPIFYWDE---------------GRSWVLTRYHDVS 59
Db         12 PFGAGEAPAY---PFHAPDRLE---PDPYWEPLRRERPLQRVTLPYGGEAWLATRYQDVR 65

Qy         60 AVFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPE--------DHARVRKLVN 111
Db         66 AVFADRRFSRQLAVAPGAPRFLPHQP-------------PPDAVLSVEGPDHARLRRLVG 112

Qy        112 PSFTSRAIDLLRAEIQRTVDQLLDARSGQEEF----DVVRDYAEGIPMRAISALLKVPAE 167
Db        113 KVFTPRRVEDMRPLIQRTADGLLDA---MEEMGPPADLVEDFSLPFAVSMICELLGVPPE 169

Qy        168 CDEKFRRFGSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLD----ERRRNPLEN 223
Db        170 DRKRFCVWSDALLTT------------TAHTPAQVRDYMMQMHDYLGGLVAQRRVRPTA- 216

Qy        224 DVLTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAE 283
Db        217 DLIGSLVTARDEEDKLTEGELVRLAEAILIAGYETSASQIPNFLYVLFRHPQLLERIRND 276

Qy        284 PGLMRNALDEVLRFDNILRIGTV----RFARQDLEYCGASIKKGEMVFLLIPSALRDGTV 339
Db        277 HDLIPDAVEELLRF---VPIGTVDGFPRTATEDVELGGVLVRAGETVVPSMGAANRDPEL 333

Qy        340 FSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKL 391
Db        334 FTDPDELDLARRPNPHLGFGAGPHHCLGAQLARVELQITLTTLFRRYPRLRL 385


RESULT 13
Q8UFY8
ID   Q8UFY8      PRELIMINARY;      PRT;   415 AA.
AC   Q8UFY8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cytochrome P450 hydroxylase.
GN   ATU1256 OR AGR_C_2319.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009087; AAL42265.1; -.
DR   EMBL; AE008053; AAK87051.1; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Complete proteome.
SQ   SEQUENCE   415 AA;  47458 MW;  920520E0C6C95DDA CRC64;

  Query Match             21.2%;  Score 451;  DB 16;  Length 415;
  Best Local Similarity   29.1%;  Pred. No. 2.6e-24;
  Matches  118;  Conservative   82;  Mismatches  175;  Indels   30;  Gaps    9;

Qy         22 PGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERF------AVSREEWE 75
Db         24 PAFYNDPNPVYAALHAQCPTFYWEEQRQWFFTCYDHVSTLLRDRRFGRQILHVASREE-- 81

Qy         76 SSAEYSSAIPE-LSDMK------KYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQR 128
Db         82 ------IGLPEPLEHVKHFDLAEQHSLLELEPPEHTRLRTLINRAFVSRHVDKMKPEIEE 135

Qy        129 TVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLV 188
Db        136 LANRLIEAFEANGETELLSSYADIIPVTMIARMIGIPEEMGPQLLKWSHA---YVGMYMF 192

Qy        189 PQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALV 248
Db        193 KRTPEDELLADKAAQEFSDYVRRVIAERRAEP-KDDLLSHMIHTEHKGQYLTDDELVSTT 251

Qy        249 GAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRF 308
Db        252 IVLLNAGHEATVHQIGNSVRIILESGLDPKTLFHDETATERTVEETLRICAPVHI-FQRW 310

Qy        309 ARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGV 368
Db        311 VLEPVEIDGVQFKRGDKVSLILAAANLDPAKFSDPLAFQPDRNEGANVSFGAGIHFCIGA 370

Qy        369 SLARLEAEIAVGTIFRRFPEMKLKETPVFG--YHPAFRNIESLNV 411
Db        371 PLARLELNLALPLLFKRLPGLKIAEPPKVKDVYH--FHGLERLDL 413


RESULT 14
Q9KHJ7
ID   Q9KHJ7      PRELIMINARY;      PRT;   401 AA.
AC   Q9KHJ7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative cytochrome P450 107R1 (Putative cytochrome P450 monooxygenase
DE   EncR).
GN   CYP107R1 OR ENCR.
OS   Streptomyces maritimus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=115828;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RA   Piel J., Hoang K., Moore B.S.;
RT   "Natural Metabolic Diversity Encoded by the Enterocin Biosynthesis
RT   Gene Cluster.";
RL   J. Am. Chem. Soc. 122:5415-5416(2000).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF254925; AAF81737.1; -.
DR   HSSP; Q00441; 1OXA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   401 AA;  44184 MW;  73CDA22BD475B2E7 CRC64;

  Query Match             21.1%;  Score 447.5;  DB 2;  Length 401;
  Best Local Similarity   31.9%;  Pred. No. 4.4e-24;
  Matches  126;  Conservative   70;  Mismatches  166;  Indels   33;  Gaps   12;

Qy          9 SETKPAFDFKPFAPGYAEDPFPAIERLREATPIF-----YWDEGRSWVLTRYHDVSAVFR 63
                : |:   || ||||           ||  :|||  ||       |    | :|::|||  |:  |
Db          3 THTQQLRDF-PFAPPAELHMEPAFAQLREEEPISRVRLPY--GGEAWLVTRYQDIKTVLG 59

Qy         64 DERFAVSREEWESSAEYSSAIPELSDMKKYG-LFGLPPEDHARVRKLVNPSFTSRAIDLL 122
                | ||:        :: :::  |       |      |  |   |  |  ||  |:||  |      ||  |  :: |
Db         60 DPRFS------RAATQHAQAPRIQPDPAGEGVLMSLDPPDHTRLRKTVAGVFTKRRVEDL 113

Qy        123 RAEIQRTVDQLLDA-RSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATAR 181
                |    ||  ::||:|   :        |:|    ||    :|:      ||  ||  :   |:  :   |
Db        114 RPATQRIAEELLEAMEASGAPADLVASYALPLPVTVICDLLGVPGDDREQLRGWSDAL-- 171

Qy        182 ALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQA-EADGSRLS 240
                  :               |:    :: :       |  |     :  :|||  |  :  |:|    |:|     :  :
Db        172 ---LSTTACTPAESAAAAQAMADHFAAL---VSQRRRQPTD-DLLGALVQTWDREEGLLR 224

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
                :|||  |      ::  ||  :||      ||          ||:  |    :: :: :  :|     |  :|::|:|||
Db        225 DEELVLLTRDLLIAGHETTASQIANCTYLLLQRPHDMDRLRTDPSAMASAVEELLRF--- 281

Qy        301 LRIGT----VRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASL 356
                : :|:       |  |  :  :|  ||    |:  |:  ||         :|    |    ||:  |     |:  |          :
Db        282 IPLGSGSFRARVATEPVELCGVRIQPGDTVFAPTVAANWDPDVFAEPGRLDIDRSPNPHV 341

Qy        357 AYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKL 391
                |:|  |  |  |  |  |   |||||  ::|:|  :  ||  |  ::|
Db        342 AFGHGVHHCLGAQLARLELQVALGVLLRRLPRLRL 376


RESULT 15
Q8KNC8
ID   Q8KNC8      PRELIMINARY;      PRT;   394 AA.
AC   Q8KNC8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CalE10.
GN   CALE10.
OS   Micromonospora echinospora (Micromonospora purpurea).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Micromonosporineae; Micromonosporaceae; Micromonospora.
OX   NCBI_TaxID=1877;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 15839;
RA   Ahlert J., Shepard E., Lomovskaya N., Zazopoulos E., Staffa A.,
RA   Bachmann B.O., Huang K., Fonstein L., Czisny A., Whitwam R.E.,
RA   Farnet C.M., Thorson J.S.;
RT   "The calicheamicin gene cluster and its iterative type I PKS.";
RL   Science 0:0-0(2002).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; AF497482; AAM94800.1; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Heme; Monooxygenase; Oxidoreductase.
SQ   SEQUENCE   394 AA;  43871 MW;  AA79A2518A2EFBDC CRC64;

  Query Match             21.0%;  Score 446.5;  DB 2;  Length 394;
  Best Local Similarity   32.1%;  Pred. No. 5.1e-24;
  Matches  131;  Conservative   64;  Mismatches  162;  Indels   51;  Gaps   13;

Qy         19 PFAPGYAEDPFPA-IERLREATPIF---YWDEGRSWVLTRYHDVSAVFRDERFAVSRE-- 72
                || |          ||    |||     |:               :||::|| ||    |    | |   :    |
Db          6 PFGP-------PAEYARLRTERPVARLPMLGGNTAWVVSRYADVKRVLSDPRMSADRRRA 58

Qy         73 ---------EWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLR 123
                         | :   | :::   | |  |:  |              |||  |      |:  :      |  :|  :   ||
Db         59 GFPRFAPTTESQRQASFANFRPPLNWMD-------PPE-HTAARRQIVDEFAARRVRQLR 110

Qy        124 AEIQRTVDQLLDA-RSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARA 182
                ::|  ||:  |||    :|:        |:|    ::    :|   |   |    :|   ||               ||           :  |
Db        111 PLVERVVDEHLDAMTAGRSSADLVPSFSYPVPSRVICEMLGVPYGEHAFFER---RSTRM 167

Qy        183 LGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTK 242
                |    |  ||   ||     :                   |         |    |:  ::  |:|    :|||:  ||  |:                      :
Db        168 LSRG-VP-ADERAR----CAREIREFLDGVVTDKERHP-GDDVLSRLLAAQRAAGEPDHE 220

Qy        243 ELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILR 302
                :|::     ::  ||   ||    :|:  :||  ||    |||    ||    ::|||          |::|:||:    |:
Db        221 AVVSMAFVLLVAGHVTTSNMISLSVLALLTHPERLARLRAEPDRFPAAVEELLRYFTIVE 280

Qy        303 IGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGP 362
                |    |    |   |:       |  :|:  ||   |     |     :|  ||      | ·|||  ||     ||           ||:|  |
Db        281 AATARTATADVTVGGVTIRAGEGVVALGQAANRDPAAFDRPDEFDPDRDARHHLAFGYGR 340

Qy        363 HVCPGVSLARLEAEIAVGTIFRRFPEMKL----------KETPVFGYH 400
                |:|||    |||||  ::|:    :   ||   |   ::|                    ::    :||  |
Db        341 HICPGQHLARLELDVALSRLVRRLPGLRLTVDVDDLPLKEDGNIFGLH 388

Search completed: October  2, 2003, 17:31:48
Job time : 41.1179 secs
```

```
OM protein - protein search, using sw model

Run on:          October  2, 2003, 17:27:29 ; Search time 7.74239 Seconds
                                              (without alignments)
                                              2544.975 Million cell updates/sec

Title:           US-10-014-717-8
Perfect score:   2124
Sequence:        1 MTQEQANQSETKPAFDFKPF..........HPAFRNIESLNVILKPSKAG 419

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:      127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_41:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	531	25.0	405	1	CPXK_SACER	P33271 saccharopol
2	525.5	24.7	395	1	BIOI_BACSU	P53554 bacillus su
3	518.5	24.4	396	1	YJIB_BACSU	O34374 bacillus su
4	502	23.6	410	1	CPXY_BACSU	O08469 bacillus su
5	459	21.6	402	1	CPXF_STRGO	P18327 streptomyce
6	450	21.2	402	1	C123_MYCTU	P77902 mycobacteri
7	434.5	20.5	438	1	C140_MYCTU	O08464 mycobacteri
8	420	19.8	405	1	CPXM_BACSU	P27632 bacillus su
9	415.5	19.6	402	1	C13B_XYLFA	Q9pgc5 xylella fas
10	411	19.4	404	1	CPXJ_SACER	Q00441 saccharopol
11	386.5	18.2	408	1	NOR2_CYLTO	Q12599 cylindrocar
12	386	18.2	398	1	C142_MYCTU	O53563 mycobacteri
13	378	17.8	381	1	CPXG_STRSQ	P23296 streptomyce
14	376.5	17.7	403	1	NOR1_CYLTO	Q00616 cylindrocar
15	373.5	17.6	489	1	C128_MYCTU	Q59572 mycobacteri
16	367.5	17.3	400	1	CPXP_RHISN	P55544 rhizobium s
17	366	17.2	410	1	CPXM_BACME	Q06069 bacillus me
18	364.5	17.2	399	1	FAS1_RHOFA	P46373 rhodococcus
19	361.5	17.0	412	1	CPXH_STRGR	P26911 streptomyce
20	359	16.9	401	1	CPXP_BRAJA	Q59203 bradyrhizob
21	357.5	16.8	409	1	CPS2_STRCC	Q59831 streptomyce
22	356.5	16.8	400	1	C13C_XYLFA	Q9pge6 xylella fas
23	354.5	16.7	405	1	CPXE_STRGO	P18326 streptomyce
24	353.5	16.6	402	1	NOR_FUSOX	P23295 fusarium ox
25	352.5	16.6	405	1	CYPX_BACSU	O34926 bacillus su
26	349.5	16.5	414	1	Y4VG_RHISN	Q53215 rhizobium s
27	346.5	16.3	429	1	CPXR_BRAJA	Q59204 bradyrhizob
28	344.5	16.2	400	1	C141_MYCTU	O08362 mycobacteri
29	344.5	16.2	433	1	C125_MYCTU	P71856 mycobacteri
30	342.5	16.1	428	1	CPXL_PSESP	P33006 pseudomonas
31	332	15.6	410	1	CPXI_BACME	P14762 bacillus me
32	328.5	15.5	406	1	CPXO_PSEPU	Q59723 pseudomonas
33	328.5	15.5	467	1	CPXR_RHISN	P55543 rhizobium s
34	326.5	15.4	428	1	C124_MYCTU	Q50696 mycobacteri
35	311	14.6	436	1	THCB_RHOER	P43492 rhodococcus
36	309	14.5	397	1	CPXQ_SACER	P48635 saccharopol
37	305	14.4	414	1	C126_MYCTU	P77903 mycobacteri
38	302.5	14.2	393	1	C143_MYCTU	O53936 mycobacteri
39	286	13.5	422	1	CPXC_AGRTU	P24466 agrobacteri
40	285.5	13.4	405	1	C130_MYCTU	Q11062 mycobacteri
41	281.5	13.3	368	1	CPXW_SULSO	Q55080 sulfolobus
42	271.5	12.8	414	1	CPXA_PSEPU	P00183 pseudomonas
43	252	11.9	407	1	CPXD_AGRTU	P24467 agrobacteri
44	242.5	11.4	434	1	C144_MYCTU	O33180 mycobacteri
45	241.5	11.4	396	1	C121_MYCTU	Q59571 mycobacteri

ALIGNMENTS

```
RESULT 1
CPXK_SACER
ID   CPXK_SACER     STANDARD;      PRT;   405 AA.
AC   P33271;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 107B1 (EC 1.14.-.-) (P450CVIIB1).
GN   CYP107B1.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 172-203 AND 302-321.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92121109; PubMed=1732208;
RA   Andersen J.F., Hutchinson C.R.;
RT   "Characterization of Saccharopolyspora erythraea cytochrome P-450
RT   genes and enzymes, including 6-deoxyerythronolide B hydroxylase.";
RL   J. Bacteriol. 174:725-735(1992).
CC   -!- FUNCTION: NOT KNOWN, PROBABLY INVOLVED IN THE CATABOLISM OF
CC       OCTANE AND GUAIACOL. IT DISPLAYS A WEAK ACTIVITY IN THE O-
CC       DEALKYLATION OF 7-ETHOXYCOUMARIN.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M83110; AAA26483.1; -.
DR   PIR; B42606; B42606.
DR   HSSP; Q00441; 1OXA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT   METAL       352    352       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   405 AA;  45238 MW;  71C93CEC1FDC53FD CRC64;

  Query Match             25.0%;  Score 531;  DB 1;  Length 405;
  Best Local Similarity   32.5%;  Pred. No. 3.1e-26;
  Matches 132;  Conservative   83;  Mismatches  169;  Indels   22;  Gaps    6;

Qy         16 DFKPFAPGYAEDPFPAIERLRE--ATPIFYWDEGRSWVLTRYHDVSAVFRDER----FAV 69
               |    |   :|:|       |:||     |    :   :|::||| ||      | |    |
Db          8 DLLAFDDAFAQDRHNRYARMREEPVQRIRTVNGLDAWLITRYEDVKQALLDPRIAKDFGR 67

Qy         70 SREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRT 129
              :::  |     :   |  |      :      | || |:||||  :||:| :: ||  |::
Db         68 TQQIIEKRLADAERRPGFSPDLGPHMLNTDPPDHTRLRKLVVKAFTARRVEGLRPRIEQI 127
```

```
Qy        130 VDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVP 189
                |||  :|: | |:: ::| :|:  || |: |     : ||  :          |   |   |
Db        128 TDDLLDRLAGRSEVDLIDEFAFPLPITVISELMGVEDSRRDDFRSW----TNVLVDGSQP 183

Qy        190 QVDEETK-TLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALV 248
                :   :    :|  :||·:|           ::|    :|:|| ||:|  || |||  ||:|:|
Db        184 EAQAQASVAMVEYLTELIA--------KKRTEPGDDLLTALLEAVEDGDRLSEGELIAMV 235

Qy        249 GAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRF 308
                 :: || :||: ||    ||:|| :|: |   :: :| |:   |::| ||::: :   || |
Db        236 FLLLVAGHETTVNLIGNCVLSLLGNPDQLAALRNDPSLLPGAIEETLRYESPVANGTFRH 295

Qy        309 ARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGV 368
                 : : :     | :||:|:: : :| |||   |  || ||: |:|    :|:| | | | |
Db        296 TAEAVRFGDVVIPEGELVWVALGAANRDGERFEDPDRFDITRETTGHVAFGHGIHFCVGA 355

Qy        369 SLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
               :|||||:||||  :   |||::::  :|     : :      | :| | |
Db        356 ALARLEAQIAVGRLLERFPDLRMAASPDDLRWRFSVLMRGLEKLPV 401
```

RESULT 2
BIOI_BACSU

```
ID   BIOI_BACSU     STANDARD;      PRT;   395 AA.
AC   P53554;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Biotin biosynthesis; cytochrome P450-like enzyme (EC 1.14.-.-).
GN   BIOI OR CYP107H.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96312354; PubMed=8763940;
RA   Bower S., Perkins J.B., Yocum R.R., Howitt C.L., Rahaim P.,
RA   Pero J.;
RT   "Cloning, sequencing, and characterization of the Bacillus subtilis
RT   biotin biosynthetic operon.";
RL   J. Bacteriol. 178:4122-4130(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98048467; PubMed=9387221;
RA   Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT   "Sequencing and functional annotation of the Bacillus subtilis genes
RT   in the 200 kb rrnB-dnaB region.";
RL   Microbiology 143:3431-3441(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U51868; AAB17462.1; -.
DR   EMBL; AF008220; AAC00266.1; -.
DR   EMBL; Z99119; CAB14997.1; -.
DR   PIR; G69594; G69594.
DR   HSSP; Q55080; 1IO7.
DR   SubtiList; BG11528; bioI.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Biotin biosynthesis; Oxidoreductase; Monooxygenase;
KW   Electron transport; Heme; Complete proteome.
FT   METAL       345    345       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   395 AA;  44865 MW;  E4AC3AF2637ACE1A CRC64;

  Query Match             24.7%;  Score 525.5;  DB 1;  Length 395;
  Best Local Similarity   31.9%;  Pred. No. 6.5e-26;
  Matches  129;  Conservative   98;  Mismatches  137;  Indels   41;  Gaps   12;

Qy         24 YAEDPFPAIERLREATPIFYWDEGR-----SWVLTRYHDVSAVFRDERFAVSREEWESSA 78
              : ::|:    :  ||    ||:    :|           | :|  |  : :|: :| |||  |        |||
Db         12 FLKNPYSFYDTLRAVHPIY---KGSFLKYPGWYVTGYEETAAILKDARFKVRTPLPESST 68

Qy         79 EYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARS 138
              :|·     :|| ::      :         || |:|  |  : :||  |     :    :     |    ||  |||
Db         69 KYQ----DLSHVQNQMMLFQNQPDHRRLRTLASGAFTPRTTESYQPYIIETVHHLLDQVQ 124

Qy        139 GQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTL 198
              |::: :|: |·:|  :          |: :: ||  |    |:  :  : ::          |:   :|   |::
Db        125 GKKKMEVISDFAFPLASFVIANIIGVPEEDREQLKEWAAS--------LIQTID-FTRSR 175

Qy        199 VASVTEG-------LALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAI 251
               | :|||            :|     ::: :|:|:| : |:::|||:    :    :|: :|    :          :
Db        176 KA-LTEGNIMAVQAMAYFKELIQKRKRHP-QQDMISMLLKGR-EKDKLTEEEAASTCILL 232

Qy        252 IAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQ 311
                 || :||: ||: :|| ||: ||  |      ::     |  |:    |::| ||:::    :: |  | |  :
Db        233 AIAGHETTVNLISNSVLCLLQHPEQLLKLRENPDLIGTAVEECLRYESPTQM-TARVASE 291

Qy        312 DLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLA 371
              |:: || :|::|| |:|:|: :| || ::|: |||||: |          |::|  | ||| | |||
Db        292 DIDICGVTIRQGEQVLLLGAANRDPSIFTNPDVFDITRSPNPHLSFGHGHHVCLGSSLA 351

Qy        372 RLEAEIAVGTIFRRFPEMKLKE-----TPVFGYHPAFRNIESLNV 411
              ||||:||: |: :| |  : |                   |:||          || :| | |
Db        352 RLEAQIAINTLLQRMPSLNLADFEWRYRPLFG----FRALEELPV 392
```

RESULT 3
YJIB_BACSU

```
ID   YJIB_BACSU     STANDARD;      PRT;   396 AA.
```

```
AC   O34374;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative cytochrome P450 yjiB (EC 1.14.-.-).
GN   YJIB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Rivolta C., Soldo B., Lazarevic V., Joris B., Mauel C., Karamata D.;
RT   "A 35.7 kb DNA fragment from Bacillus subtilis chromosome containing a
RT   putative 12.3 kb operon involved in hexuronate catabolism and a
RT   perfect catabolite-responsive element.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF015825; AAC46317.1; -.
DR   EMBL; Z99110; CAB13078.1; -.
DR   PIR; B69851; B69851.
DR   HSSP; Q55080; 1IO9.
DR   SubtiList; BG13195; yjiB.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 2.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW   Complete proteome.
FT   METAL       349    349       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   396 AA;  44990 MW;  9A89CF12613DBCFB CRC64;

  Query Match             24.4%;  Score 518.5;  DB 1;  Length 396;
  Best Local Similarity   32.6%;  Pred. No. 1.8e-25;
  Matches  124;  Conservative   72;  Mismatches  155;  Indels   29;  Gaps    6;

Qy         25 AEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAI 84
              |  |||  | :|:  |: : :| : |  :   | ||  |  |:             :|| :
Db         23 AYHPFPWYESMRKDAPVSFDEENQVWSVFLYDDVKKVVGDKEL------------FSSCM 70

Qy         85 PELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEEFD 144
              |: :        :  : |  | ::| :|| :|| | :         ||   |:|:    |: |||
Db         71 PQQTSSIGNSIINMDPPKHTKIRSVVNKAFTPRVMKQWEPRIQEITDELIQKFQGRSEFD 130

Qy        145 VVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLVA---- 200
              :| |::  :|:  || || ||:     |:|: :             | |          || :
Db        131 LVHDFSYPLPVIVISELLGVPSAHMEQFKAWSD-----LLVSTPKDKSEEAEKAFLEERD 185

Qy        201 SVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTI 260
                   | ||        : | :|| | |::::|::|| | :|| :||:     :: || :||
Db        186 KCEEELAAFFAGIIEEKRNKPEQDIISILVEAEETGEKLSGEELIPFCTLLLVAGNETTT 245

Qy        261 YLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRF---DNILRIGTVRFARQDLEYCG 317
                   ||:  |: ::| :|   | ::: | ||  |::| |||       :||   | |::|  |  |
Db        246 NLISNAMYSILETPGVYEELRSHPELMPQAVEEALRFRAPAPVLR----RIAKRDTEIGG 301

Qy        318 ASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLEAEI 377
                   ||:|:||    : || ||     | || :||:||       :|:| | | | | |   |||||| |
Db        302 HLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHPNPHIAFGHGIHFCLGAPLARLEANI 361

Qy        378 AVGTIFRRFPEMK-LKETPV 396
              |: ::   || |: :  ||:
Db        362 ALTSLISAFPHMECVSITPI 381

RESULT 4
CPXY_BACSU
ID   CPXY_BACSU     STANDARD;      PRT;   410 AA.
AC   O08469;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 (EC 1.14.-.-).
GN   CYPA OR CYP107J1.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE=97431495; PubMed=9287000;
RA   Belitsky B.R., Gustafsson M.C.U., Sonenshein A.L., von Wachenfeldt C.;
RT   "An lrp-like gene of Bacillus subtilis involved in branched-chain
RT   amino acid transport.";
RL   J. Bacteriol. 179:5448-5457(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97453479; PubMed=9308178;
RA   Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,
RA   Duesterhoeft A., Ehrlich S.D.;
RT   "Sequence of the Bacillus subtilis genome region in the vicinity of
RT   the lev operon reveals two new extracytoplasmic function RNA
RT   polymerase sigma factors SigV and SigZ.";
RL   Microbiology 143:2939-2943(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
```

```
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y11043; CAA71937.1; -.
DR   EMBL; U93876; AAB80898.1; -.
DR   EMBL; Z99117; CAB14615.1; -.
DR   PIR; E69611; E69611.
DR   HSSP; Q00441; 1OXA.
DR   SubtiList; BG11929; cypA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Heme; Complete proteome.
FT   METAL       359    359       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   410 AA;  47384 MW;  035E98E58A4761AB CRC64;

  Query Match             23.6%;  Score 502;  DB 1;  Length 410;
  Best Local Similarity   30.9%;  Pred. No. 2e-24;
  Matches  115;  Conservative   88;  Mismatches  159;  Indels   10;  Gaps    4;

Qy         24 YAEDPFPAIERLREAT---PIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAEY 80

Db         22 FKDEAYEFYKELRKSQALYPLSLGALGKGWLISRYDDAIHLLKNEKLKKNYENVFTAKEK 81

Qy         81 SSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQ 140

Db         82 RPALLKNEETLTKHMLNSDPPDHNRLRTLVQKAFTHRMILQLEDKIQHIADSLLDKVQPN 141

Qy        141 EEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLVA 200

Db        142 KFMNLVDDYAFPLPIIVISEMLGIPLEDRQKFRVWSQA---IIDFSDAPERLQENDHLLG 198

Qy        201 SVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTI 260

Db        199 EFVE---YLESLVRKKRREP-AGDLISALIQAESEGTQLSTEELYSMIMLLIVAGHETTV 254

Qy        261 YLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEYCGASI 320

Db        255 NLITNMTYALMCHHDQLEKLRQQPDLMNSAIEEALRFHSPVELTTIRWTAEPFILHGQEI 314

Qy        321 KKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLEAEIAVG 380

Db        315 KRKDVIIISLASANRDEKIFPNADIFDIERKNNRHIAFGHGNHFCLGAQLARLEAKIAIS 374

Qy        381 TIFRRFPEMKLK 392

Db        375 TLLRRCPNIQLK 386

RESULT 5
CPXF_STRGO
ID   CPXF_STRGO     STANDARD;      PRT;   402 AA.
AC   P18327;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450-SU2 (EC 1.14.-.-) (P450-CVB1) (CYP105B1).
GN   CYP105B1 OR SUBC.
OS   Streptomyces griseolus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1909;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-32.
RC   STRAIN=ATCC 11796;
RX   MEDLINE=90264332; PubMed=2345149;
RA   Omer C.A., Lenstra R., Litle P.J., Dean C., Tepperman J.M.,
RA   Leto K.J., Romesser J.A., O'Keefe D.P.;
RT   "Genes for two herbicide-inducible cytochromes P-450 from
RT   Streptomyces griseolus.";
RL   J. Bacteriol. 172:3335-3345(1990).
CC   -!- FUNCTION: METABOLISM OF A NUMBER OF SULFONYLUREA HERBICIDES.
CC   -!- INDUCTION: By herbicides.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M32239; AAA26825.1; -.
DR   PIR; B35401; B35401.
DR   HSSP; P23295; 1CMN.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Heme.
FT   INIT_MET      0      0
FT   METAL       351    351       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   402 AA;  44278 MW;  E3B67F6F72C26D9D CRC64;

  Query Match             21.6%;  Score 459;  DB 1;  Length 402;
  Best Local Similarity   31.7%;  Pred. No. 9.3e-22;
  Matches  127;  Conservative   73;  Mismatches  157;  Indels   44;  Gaps   12;

Qy         21 APGYAEDPFPAIERLREATPI---FYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESS 77

Db         19 APGCPFDPAPDVTEAARTEPVTRATLWDGSSCWLVTRHQDVRAVLGDPRF--------SA 70

Qy         78 AEYSSAIPELSDMKKYGLFGLPP-------EDHARVRKLVNPSFTSRAIDLLRAEIQRTV 130

Db         71 DAHRTGFPFLTAGGR-EIIGTNPTFLRMDDPEHARLRRMLTADFIVKKVEAMRPEVQRLA 129

Qy        131 DQLLD-ARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVP 189

Db        130 DDLVDRMTTGRTSADLVTEFALPLPSLVICLLLGVPYE-DHAF--FQERSRVLLTLRSTP 186
```

```
Qy        190 QVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVG 249
                  || :     : | || |        :|  | :  :  :: : | | ::::|   : ||
Db        187 ---EEVRAAQDELLEYLARL---ARTKRERPDDAIISRLVARGELDDTQIATMGRLLLV- 239

Qy        250 AIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFA 309
                  || :||   : | : | |||:|: |   ::||| |:: |::|:||:  |:  |   | |
Db        240 ----AGHETTANMTALSTLVLLRNPDQLARLRAEPALVKGAVEELLRYLTIVHNGVPRIA 295

Qy        310 RQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVS 369
                 :|:   | :|  || |  :| || ||  ||   |   | ||     :|:| | | | |
Db        296 TEDVLIGGRTIAAGEGVLCMISSANRDAEVFPGGDDLDVARDARRHVAFGFGVHQCLGQP 355

Qy        370 LARLEAEIAVGTIFRRFPEMKL----KETP------VFGYH 400
              |||:| :||: |: || |::: |     :| |        ::| |
Db        356 LARVELQIAIETLLRRLPDLRLAVPHEEIPFRGDMAIYGVH 396
```

RESULT 6
C123_MYCTU
ID C123_MYCTU STANDARD; PRT; 402 AA.
AC P77902;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative cytochrome P450 123 (EC 1.14.-.-).
GN CYP123 OR RV0766C OR MT0790 OR MTCY369.11C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC --

CC --
DR EMBL; Z80226; CAB02396.1; -.
DR EMBL; AE006970; AAK45032.1; -.
DR PIR; A70707; A70707.
DR HSSP; Q00441; 1OXA.
DR TIGR; MT0790; -.
DR TubercuList; Rv0766c; -.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 350 350 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 402 AA; 45421 MW; 76AD843019361798 CRC64;

Query Match 21.2%; Score 450; DB 1; Length 402;
Best Local Similarity 32.4%; Pred. No. 3.4e-21;
Matches 128; Conservative 64; Mismatches 179; Indels 24; Gaps 9;

```
Qy         13 PAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRD-----ERF 67
                  |:   : |||:|   |||:  |::  :|   | ::|:|||   |||          :
Db          8 PELVLDPYDYDFHEDPYPYYRRLRDEAPLYRNEERNFWAVSRHHDVLQGFRDSTALSNAY 67

Qy         68 AVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQ 127
              ||    : |:  | |   :|          :   :     | |:| ||:  || | |   ::
Db         68 GVS---LDPSSRTSEAYRVMS------MLAMDDPAHLRMRTLVSKGFTPRRIRELEPQVL 118

Qy        128 RTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA-RALGVG 186
                        ||:    | || | ::|  :||  || |: ||      : |     |    | ||
Db        119 ELARIHLDSALQTESFDFVAEFAGKLPMDVISELIGVPDTDRARIRALADAVLHREDGVA 178

Qy        187 LVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVA 246
                ||         : ||: | :     |:: | || |  |:: : || || || ||| :|::|
Db        179 DVP-----PPAMAASI-ELMRYYADLIAEFRRRP-ANNLTSALLAAELDGDRLSDQEIMA 231

Qy        247 LVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTV 306
                :  :: || :||  |:| ||           |   |  | |:   :    ::| ||:|    :|
Db        232 FLFLMVIAGNETTTKLLANAVYWAAHHPGQLARVFADHSRIPMWVEETLRYDTSSQI-LA 290

Qy        307 RFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASL-AYGRGPHVC 365
                  |  · ||      :| :||:: ||  ||   ||  ||  || : :  |: |    | ::| | | |
Db        291 RTVAHDLTLYDTTIPEGEVLLLLPGSANRDDRVFDDPDDYRIGREIGCKLVSFGSGAHFC 350

Qy        366 PGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYH 400
                |  |||:|| :|:| : ||     :: :   |    |
Db        351 LGAHLARMEARVALGALLRRIRNYEVDDDNVVRVH 385
```

RESULT 7
C140_MYCTU
ID C140_MYCTU STANDARD; PRT; 438 AA.
AC O08464;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative cytochrome P450 140 (EC 1.14.-.-).
GN CYP140 OR RV1880C OR MT1929 OR MTCY180.38.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;

```
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z97193; CAB10066.1; -.
DR   EMBL; AE007049; AAK46201.1; ALT_INIT.
DR   PIR; E70515; E70515.
DR   HSSP; Q00441; 1OXA.
DR   TIGR; MT1929; -.
DR   TubercuList; Rv1880c; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW   Complete proteome.
FT   METAL       381    381       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   438 AA;  48871 MW;  0463B3EECF358744 CRC64;

  Query Match             20.5%;  Score 434.5;  DB 1;  Length 438;
  Best Local Similarity   29.6%;  Pred. No. 3.5e-20;
  Matches  126;  Conservative   77;  Mismatches  168;  Indels   55;  Gaps   13;

Qy         22 PGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVS-AVFRDERFAVSREEWESSAEY 80

Db         35 PAVATDPVPFYDEVRSHGALV---RNRANYLTVDHRLAHDLLRSDDFRV--------VSF 83

Qy         81 SSAIP-------------ELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQ 127

Db         84 GENLPPPLRWLERRTRGDQLHPLREPSLLAVEPPDHTRYRKTVSAVFTSRAVSALRDLVE 143

Qy        128 RTVDQLLDARSGQEEF-DVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVG 186

Db        144 QTAINLLDRFAEQPGIVDVVGRYCSQLPIVVISEILGVPEHDRPRVLEFGELAAPSLDIG 203

Qy        187 LVPQVDEETKTLVASVTEGL----ALLHDVLDERRRNPLENDVLTMLLQAEADG---SRL 239

Db        204 I-------PWRQYLRVQQGIRGFDCWLEGHLQQLRHAP-GDDLMSQLIQIAESGDNETQL 255

Qy        240 STKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDN 299

Db        256 DETELRAIAGLVLVAGFETTVNLLGNGIRMLLDTPEHLATLRQHPELWPNTVEEILRLDS 315

Qy        300 ILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRR-DTGASLAY 358

Db        316 PVQL-TARVACRDVEVAGVRIKRGEVVVIYLAAANRDPAVFPDPHRFDIERPNAGRHLAF 374

Qy        359 GRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKL------KETPVFGYHPAFRNIESLNVI 412

Db        375 STGRHFCLGAALARAEGEVGLRTFFDRFPDVRAAGAGSRRDTRV------LRGWSTLPVT 428

Qy        413 LKPSKA 418

Db        429 LGPARS 434


RESULT 8
CPXM_BACSU
ID   CPXM_BACSU     STANDARD;      PRT;   405 AA.
AC   P27632;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 109 (EC 1.14.-.-) (ORF405).
GN   CYP109.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W23;
RX   MEDLINE=91192601; PubMed=1849493;
RA   Ahn K.S., Wake R.G.;
RT   "Variations and coding features of the sequence spanning the
RT   replication terminus of Bacillus subtilis 168 and W23 chromosomes.";
RL   Gene 98:107-112(1991).
CC   -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC       MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC       COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M24523; AAA22720.1; -.
DR   HSSP; Q55080; 1IO7.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Heme.
FT   METAL       351    351       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   405 AA;  45845 MW;  1331D5BEA74E3C04 CRC64;

  Query Match             19.8%;  Score 420;  DB 1;  Length 405;
  Best Local Similarity   29.0%;  Pred. No. 2.5e-19;
  Matches  121;  Conservative   72;  Mismatches  182;  Indels   42;  Gaps    9;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAED----PFPAIERLREATPIFYWDEGRSWVLTRYH 56

Db          1 MTNQTARSSKKERYANLIPMEELHSEKDRLFPFPIYDKLRRESPVRYDPLRDCWDVFKYD 60

Qy         57 DVSAVFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTS 116

Db         61 DVQFVLKNPKLFSSKRGIQTES----------------ILTMDPPKHTKLRALVSRAFTP 104

Qy        117 RAIDLLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFG 176

Db        105 KAVKQLETRIKDVTAFLLQEARQKSTIDIIEDFAGPLPVIIIAEMLGAPIEDRHLIKTYS 164

Qy        177 SATARALGVGLVPQVDEETKTLVASVTEGLALL----HDVLDERRRNPLENDVLTMLLQA 232

Db        165 D----VLVAGAKDSSDKAVADMVHNRRDGHAFLSDYFRDILSKRRAEPKE-DLMTMLLQA 219

Qy        233 EADGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALD 292

Db        220 EIDGEYLTEEQLIGFCILLLVAGNETTTNLIANAVRYLTEDSVVQQQVRQNTDNVANVIE 279

Qy        293 EVLR-FDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRD 351

Db        280 ETLRYYSPVQAIG--RVATEDTELGGVFIKKGSSVISWIASANRDEDKFCKPDCFKIDRP 337

Qy        352 TGASLAYGRGPHVCPGVSLARLEAEIAVGTIF-------RRFPEMKLKETP---VFG 398

Db        338 SYPHLSFGFGIHFCLGAPLARLEANIALSSLLSMSACIEKAAHDEKLEAIPSPFVFG 394
```

```
RESULT 9
C13B_XYLFA
ID   C13B_XYLFA     STANDARD;      PRT;   402 AA.
AC   Q9PGC5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative cytochrome P450 133B1 (EC 1.14.-.-).
GN   CYP133B1 OR XF0377.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE003889; AAF83187.1; -.
DR   PIR; H82813; H82813.
DR   HSSP; Q00441; 1OXA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW   Complete proteome.
FT   METAL       348    348       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   402 AA;  45023 MW;  5891FDE7EC128F63 CRC64;

  Query Match             19.6%;  Score 415.5;  DB 1;  Length 402;
  Best Local Similarity   29.6%;  Pred. No. 4.8e-19;
  Matches  120;  Conservative   78;  Mismatches  176;  Indels   31;  Gaps   11;

Qy         22 PGYAEDPFPAIERLREATPIFYWDEGRSWVLT-RYHDVSAVFRDERFAVSREEWES-SAE 79
              | : |:|:|  | ||    |  :     | : ::| ||  | ::  :     : ::  ||
Db          9 PAFLENPYPLYETLRAQAP--FVSIGPNALMTGRYSLVDSLLHNRN--MGKKYMESMRVR 64

Qy         80 YSSAIPELSDMKKYGLFG-----LPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLL 134
              | :    :||  :   |         : |   |   :|  ||   :||  |   : :|       |  ||:
Db         65 YGDS---AADMPLFQAFSRMFITINPPAHTHLRGLVMQAFTGRESESMRPLAIDTAHQLI 121

Qy        135 DARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDE- 193
              |     :     |:| ::|     ||: |   :: |                ||     :  |    :  :|  :
Db        122 DNFEQKPSVDLVAEFAFPFPMQIICKMMDVD---------IGDAVTLGIAVSKIAKVFDP 172

Qy        194 ---ETKTLVASVT--EGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALV 248
                     || :  |   | ||       | | ||    |:::| |:|| || :|:  |:|: |
Db        173 SPMSADELVHASTAYEELAQYFTKLIELRRTHPGTDLISMFLRAEEDGEKLTHDEIVSNV 232

Qy        249 GAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRF 308
                 :: || :|| :|   |:: | | || | |:|::  || |: | ||:| ::   |:|
Db        233 IMLLIAGYETTSNMIGNALIALHRHPEQLALLKSDLSLMPQAVSECLRYDGSVQF-TMRA 291

Qy        309 ARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGV 368
              |   |:|   |   : :| :|||:: :| ||     |: ||   |: |    |      ::| | | | |
Db        292 AMDDIEVEGELVPRGTVVFLMLGAANRDPAQFTHPDQLDITRKQGRLQSFGAGIHHCLGY 351

Qy        369 SLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVIL 413
               || :| | |: |:| | | ::|          ::    |:  :| ::
Db        352 RLALIELECALTTLFERLPHLRLAHLDALNWNQR-SNLRGVNTLI 395


RESULT 10
CPXJ_SACER
ID   CPXJ_SACER     STANDARD;      PRT;   404 AA.
AC   Q00441;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   6-deoxyerythronolide B hydroxylase (EC 1.-.-.-) (6-DEB hydroxylase)
DE   (Erythomycin A biosynthesis hydrolase) (Cytochrome P450 107A1)
DE   (CYPCVIIA1) (P450eryF).
GN   ERYF OR CYP107A1.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92079886; PubMed=1840640;
RA   Haydock S.F., Dowson J.A., Dhillon N., Roberts G.A., Cortes J.,
RA   Leadlay P.F.;
RT   "Cloning and sequence analysis of genes involved in erythromycin
RT   biosynthesis in Saccharopolyspora erythraea: sequence similarities
RT   between EryG and a family of S-adenosylmethionine-dependent
RT   methyltransferases.";
RL   Mol. Gen. Genet. 230:120-128(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91188269; PubMed=2011746;
RA   Weber J.M., Leung J.O., Swanson S.J., Idler K.B., McAlpine J.B.;
RT   "An erythromycin derivative produced by targeted gene disruption in
RT   Saccharopolyspora erythraea.";
RL   Science 252:114-117(1991).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX   MEDLINE=95269057; PubMed=7749919;
RA   Cupp-Vickery J.L., Poulos T.L.;
RT   "Structure of cytochrome P450eryF involved in erythromycin
RT   biosynthesis.";
RL   Nat. Struct. Biol. 2:144-153(1995).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX   MEDLINE=20202586; PubMed=10716705;
RA   Cupp-Vickery J., Anderson R., Hatziris Z.;
RT   "Crystal structures of ligand complexes of P450eryF exhibiting
RT   homotropic cooperativity.";
```

```
RL   Proc. Natl. Acad. Sci. U.S.A. 97:3050-3055(2000).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE=21363699; PubMed=11469860;
RA   Cupp-Vickery J., Garcia C., Hofacre A., McGee-Estrada K.;
RT   "Ketoconazole-induced conformational changes in the active site of
RT   cytochrome P450eryF.";
RL   J. Mol. Biol. 311:101-110(2001).
CC   -!- FUNCTION: CONVERSION OF 6-DEOXYERYTHRONOLIDE B (6-DEB) TO
CC       ERYTHRONOLIDE B (EB) BY THE INSERTION OF AN OXYGEN AT THE 6S
CC       POSITION OF 6-DEB.
CC   -!- PATHWAY: Erythromycin A biosynthesis.
CC   -!- MISCELLANEOUS: A MUTATED FORM OF THIS ENZYME LEADS TO THE
CC       FORMATION OF DEOXYERYTHROMYCIN A, AN ANTIBIOTIC WITH IMPROVED
CC       PHARMACOLOGICAL PROPERTIES.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X60379; CAA42927.1; -.
DR   EMBL; M54983; AAA26496.1; -.
DR   PIR; S18531; S18531.
DR   PDB; 1OXA; 07-DEC-95.
DR   PDB; 1EGY; 02-APR-00.
DR   PDB; 1EUP; 28-APR-00.
DR   PDB; 1JIN; 17-OCT-01.
DR   PDB; 1JIO; 17-OCT-01.
DR   PDB; 1JIP; 17-OCT-01.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Antibiotic biosynthesis; 3D-structure.
FT   METAL       351    351       IRON (HEME AXIAL LIGAND).
FT   CONFLICT    263    263       Q -> QDQ (IN REF. 1).
FT   CONFLICT    302    302       G -> R (IN REF. 1).
FT   STRAND        6      6
FT   TURN          7      8
FT   HELIX        10     12
FT   TURN         13     13
FT   HELIX        16     26
FT   STRAND       29     34
FT   TURN         35     36
FT   STRAND       37     42
FT   HELIX        45     53
FT   TURN         55     56
FT   STRAND       57     58
FT   TURN         61     62
FT   TURN         66     67
FT   HELIX        73     75
FT   TURN         76     77
FT   HELIX        80     86
FT   HELIX        90     92
FT   TURN         95     96
FT   HELIX        97    106
FT   TURN        107    108
FT   HELIX       111    115
FT   TURN        116    117
FT   HELIX       118    132
FT   STRAND      137    139
FT   HELIX       140    143
FT   TURN        144    146
FT   HELIX       147    157
FT   TURN        158    158
FT   TURN        161    165
FT   HELIX       166    175
FT   HELIX       179    181
FT   HELIX       182    205
FT   HELIX       211    217
FT   STRAND      219    220
FT   TURN        221    223
FT   STRAND      224    225
FT   HELIX       228    259
FT   HELIX       261    269
FT   HELIX       271    273
FT   HELIX       274    284
FT   STRAND      291    295
FT   STRAND      299    301
FT   TURN        302    303
FT   STRAND      304    306
FT   TURN        308    309
FT   STRAND      311    314
FT   HELIX       316    319
FT   TURN        320    321
FT   TURN        323    325
FT   TURN        329    330
FT   TURN        334    335
FT   TURN        339    340
FT   TURN        343    344
FT   TURN        348    349
FT   TURN        352    353
FT   HELIX       354    371
FT   TURN        373    374
FT   STRAND      376    377
FT   HELIX       381    383
FT   STRAND      386    386
FT   STRAND      395    395
FT   STRAND      399    401
SQ   SEQUENCE   404 AA;  45099 MW;  257ABEFC2D88B3FE CRC64;

  Query Match             19.4%;  Score 411;  DB 1;  Length 404;
  Best Local Similarity   30.7%;  Pred. No. 9.2e-19;
  Matches  119;  Conservative   71;  Mismatches  149;  Indels   48;  Gaps   14;

Qy         24 YAEDPFPAIERLREATPI----FYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAE 79
              |||        |||  |:      |   :  :|::| | :  |    | |  :   ::   |
Db         20 YAE-------LRETAPVTPVRFLGQD--AWLVTGYDEAKAALSDLRLSSDPKKKYPGVE 69

Qy         80 YS----SAIPELSDMKKYGLFGL---PPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQ 132
                     || |::|    :     |  | |:||||:  || | :: :|  :::   :
Db         70 VEFPAYLGFPE--DVRNYFATNMGTSDPPTHTRLRKLVSQEFTVRRVEAMRPRVEQITAE 127

Qy        133 LLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFR-RFGSATARALGVGLVPQV 191
              |||        |:|  :|  :|::  | ||       |||:|  ||  ::  |  :   |:
Db        128 LLDEVGDSGVVDIVDRFAHPLPIKVICELLGV----DEKYRGEFGRWSSEILVMD--PER 181

Qy        192 DEETKTLVASVTEGLALLHDVLD--ERRRNPLENDVLTMLLQA-EADGSRLSTKELVALV 248
               |:         |       :: :||  |||     :|:|: |::    : |  |||  || ::
Db        182 AEQRGQAAREV------VNFILDLVERRRTEPGDDLLSALIRVQDDDDGRLSADELTSIA 235

Qy        249 GAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIG---- 304
                ::  || : :: ||        ||  |:  |  ||: :|   :  ||::|:||:       |
Db        236 LVLLLAGFEASVSLIGIGTYLLLTHPDQLALVRRDPSALPNAVEEILRY-----IAPPET 290

Qy        305 TVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHV 364
              | ||| :::|  | :| :    | :    :| ||     |   ||| |||    |::|:| |
Db        291 TTRFAAEEVEIGGVAIPQYSTVLVANGAANRDPKQFPDPHRFDVTRDTRGHLSFGQGIHF 350

Qy        365 CPGVSLARLEAEIAVGTIFRRFPEMKL 391
              |  |  ||:|| |:|:   :| ||| : |
Db        351 CMGRPLAKLEGEVALRALFGRFPALSL 377


RESULT 11
NOR2_CYLTO
ID   NOR2_CYLTO     STANDARD;      PRT;   408 AA.
AC   Q12599;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
```

```
DE   Cytochrome P450 55A3 (EC 1.14.-.-) (Cytochrome P450NOR2).
GN   CYP55A3.
OS   Cylindrocarpon tonkinense (Cylindrocarpon lichenicola).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Fusarium.
OX   NCBI_TaxID=42744;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IFO 30561;
RX   MEDLINE=97163854; PubMed=9010609;
RA   Kudo T., Tomura D., Liu D.L., Dai X.Q., Shoun H.;
RT   "Two isozymes of P450nor of Cylindrocarpon tonkinense: molecular
RT   cloning of the cDNAs and genes, expressions in the yeast, and the
RT   putative NAD(P)H-binding site.";
RL   Biochimie 78:792-799(1996).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D78512; BAA11409.1; -.
DR   PIR; JC5674; JC5674.
DR   PIR; PD0007; PD0007.
DR   HSSP; P23295; 1F24.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Heme.
FT   METAL       357    357       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   408 AA;  45381 MW;  E999298B1949E7E5 CRC64;

  Query Match             18.2%;  Score 386.5;  DB 1;  Length 408;
  Best Local Similarity   28.6%;  Pred. No. 3.1e-17;
  Matches  116;  Conservative   68;  Mismatches  174;  Indels   47;  Gaps    9;

Qy          7 NQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIF---YWDEGRSWVLTRYHDVSAVFR 63
Db          6 DETTTIPRF---PFQRASAFEPPAEFARLRANEPISQVELFDGSLAWLVVKHEDVCRVAT 62

Qy         64 DERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLP------PEDHARVRKLVNPSFTSR 117
Db         63 DERLSKERTR--------LGFPELSAGGKAAAKNKPTFVDMDAPAHMNQRSMVEPFFTED 114

Qy        118 AIDLLRAEIQRTVDQLLD---ARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDE---- 170
Db        115 HVENLRPYIKETVQGLLNDMVANGCEEPVDLIEKFALPVPSYIIYTILGVPFEDLEYLTE 174

Qy        171 --KFRRFGSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTM 228
Db        175 QNAIRSNGSGTA---------------QEAAAANQQLLKYLAKLVDQRLQEP-KDDLIGR 218

Qy        229 LLQAEADGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMR 288
Db        219 LVDQQLVPGHIEKSDAVQIAFLLLVAGNATMVNMIALGVVTLMQNPSQLEELKADPTLVP 278

Qy        289 NALDEVLRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDV 348
Db        279 GFVEELCRYHTGSSMAMKRVAKEDMELGGKLIRAGEGIIASNQSANRDEDVFPNPDVFDM 338

Qy        349 RRDTGA--SLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKL 391
Db        339 HRDFDSRDGLGFGFGPHRCIAELLAKAELEIVFETLFATLPDLRV 383


RESULT 12
C142_MYCTU
ID   C142_MYCTU     STANDARD;      PRT;   398 AA.
AC   O53563;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative cytochrome P450 142 (EC 1.14.-.-).
GN   CYP142 OR RV3518C OR MT3619 OR MTV023.25C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL022022; CAA17755.1; -.
DR   EMBL; AE007165; AAK47979.1; ALT_INIT.
DR   PIR; H70807; H70807.
DR   HSSP; P33006; 1CPT.
DR   TIGR; MT3619; -.
DR   TubercuList; Rv3518c; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW   Complete proteome.
FT   METAL       340    340       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   CONFLICT      1     18       MTEAPDVDLADGNFYASR -> GWSPNRSRAPPSRARRP
FT                                (IN REF. 2).
SQ   SEQUENCE   398 AA;  44398 MW;  BCFF3C23ECB5767F CRC64;

  Query Match             18.2%;  Score 386;  DB 1;  Length 398;
  Best Local Similarity   29.3%;  Pred. No. 3.3e-17;
  Matches  119;  Conservative   65;  Mismatches  168;  Indels   54;  Gaps   11;

Qy         36 REATPIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAI-----PELSDM 90
Db         18 REARAAYRWMRAN----------QPVFRD------RNGLAAASTYQAVIDAERQPELFS- 60

Qy         91 KKYGLFGLPPED-------------HARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDAR 137
Db         61 ---NAGGIRPDQPALPMMIDMDDPAHLLRRKLVNAGFTRKRVKDKEASIAALCDTLIDAV 117
```

```
Qy        138 SGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKT 197
              : | | ||| |  :||  |  :| |  | :| | |:         |   :  :  : |
Db        118 CERGECDFVRDLAAPLPMAVIGDMLGVRPEQRDMFLRWSDDLVTFLSSHVSQEDFQITMD 177

Qy        198 LVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGTD 257
               |:  :          :  || :|  :  |::::|: :| || ||| |||      |:  | :
Db        178 AFAAYND---FTRATIAARRADPTD-DLVSVLVSSEVDGERLSDDELVMETLLILIGGDE 233

Qy        258 TTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRF----DNILRIGTVRFARQDL 313
              || : ::       |||: :  :|:: :| |:  |::|:||:       |: |: |    |
Db        234 TTRHTLSGGTEQLLRNRDQWDLLQRDPSLLPGAIEEMLRWTAPVKNMCRVLTA-----DT 288

Qy        314 EYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARL 373
              |: | :: || : ||  ||  :  || |  |||:|:  :  ||:| | | |  |  ||||
Db        289 EFHGTALCAGEKMMLLFESANFDEAVFCEPEKFDVQRNPNSHLAFGFGTHFCLGNQLARL 348

Qy        374 EAEIAVGTIFRRFPEMKL-KETPVFGYHPA--FRNIESLNVILKPS 416
              |  :    : || |:::| :  |      ||      :||: |:  |·|
Db        349 ELSLMTERVLRRLPDLRLVADDSVLPLRPANFVSGLESMPVVFTPS 394


RESULT 13
CPXG_STRSQ
ID   CPXG_STRSQ     STANDARD;      PRT;   381 AA.
AC   P23296;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cytochrome P450 105C1 (EC 1.14.-.-).
GN   CYP105C1 OR CHOP.
OS   Streptomyces sp.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1931;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90299781; PubMed=2361941;
RA   Horii M., Ishizaki T., Paik S.Y., Manome T., Murooka Y.;
RT   "An operon containing the genes for cholesterol oxidase and a
RT   cytochrome P-450-like protein from a Streptomyces sp.";
RL   J. Bacteriol. 172:3644-3653(1990).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M31939; AAA26718.1; -.
DR   HSSP; Q00441; 1OXA.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT   METAL       330    330       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   381 AA;  41703 MW;  841B959C9DDEA99C CRC64;

  Query Match             17.8%;  Score 378;  DB 1;  Length 381;
  Best Local Similarity   30.3%;  Pred. No. 9.8e-17;
  Matches  115;  Conservative   60;  Mismatches  148;  Indels   56;  Gaps   12;

Qy         28 PFPAIERLREATPI---FYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAI 84
              | . :: ||   |:     :  |     |::| |      || | || :| | |     |:
Db         19 PPAEMQALRHKAPVTRTAFADGRPGWLVTGYSAARAVLSDSRF-TARGERE-----HPAV 72

Qy         85 PELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQ-EEF 143
              |  : :               ||  | |:|:   ||:| :  |     :| | : |||
Db         73 PRAATL----------EDE-RCRRLIAGQFTARRMRQLTGRTERIVREHLDAMEHMGSPA 121
```

```
Qy        144 DVVRDYAEGIPMRAISALLKVPAECDEKFRR-------FGSATARALGVGLVPQVDEETK 196
              |:|  :|  :|    |:  || ||      | |:              || :|          ||
Db        122 DLVEHFALPVPSLVIAELLGVPPPDREHFQHDTLRWGGFGRST------------EE-- 166

Qy        197 TLVASVTEGLALLHDVLDE----RRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAII 252
                   |||      |     |      :|   |  :|:|: |: |:      |: :|| ::    ::
Db        167 -----VTEAFVSLGGQLQRLVRLKRTEP-GDDLLSGLIAAD---PALTDEELASIAFLLL 217

Qy        253 AAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQD 312
              || || : ||      || |: |   ::|:| |   :|::|:||  :::  |  | | ||
Db        218 VAGHGTTAHQIALGAFLLLEHPDQLAALRADPALTESAVEELLRHLSVVHHGPTRAALQD 277

Qy        313 LEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLAR 372
                :  |   :|  ||:| : : :| ||      | |||  || |:          ||:| |  | | |   |||
Db        278 ADIEGTPVKAGEVVVVSLGAANRDPARFERPDAVDVTREDTGHLAFGHGMHQCLGRQLAR 337

Qy        373 LEAEIAVGTIFRRFPEMKL 391
               :|  :|:  :   ||| ::|
Db        338 IELRVALTALLERFPHLRL 356


RESULT 14
NOR1_CYLTO
ID   NOR1_CYLTO     STANDARD;      PRT;   403 AA.
AC   Q00616;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Cytochrome P450 55A2 (EC 1.14.-.-) (Cytochrome P450NOR1).
GN   CYP55A2.
OS   Cylindrocarpon tonkinense (Cylindrocarpon lichenicola).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Fusarium.
OX   NCBI_TaxID=42744;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IFO 30561;
RX   MEDLINE=97163854; PubMed=9010609;
RA   Kudo T., Tomura D., Liu D.L., Dai X.Q., Shoun H.;
RT   "Two isozymes of P450nor of Cylindrocarpon tonkinense: molecular
RT   cloning of the cDNAs and genes, expressions in the yeast, and the
RT   putative NAD(P)H-binding site.";
RL   Biochimie 78:792-799(1996).
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D78511; BAA11408.1; -.
DR   HSSP; P23295; 1F24.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 2.
DR   PROSITE; PS00086; CYTOCHROME_P450; FALSE_NEG.
KW   Oxidoreductase; Monooxygenase; Heme.
FT   METAL       352    352       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   403 AA;  44273 MW;  1C2AB643FD219534 CRC64;

  Query Match             17.7%;  Score 376.5;  DB 1;  Length 403;
  Best Local Similarity   28.0%;  Pred. No. 1.3e-16;
  Matches  116;  Conservative   59;  Mismatches  184;  Indels   55;  Gaps   10;

Qy         13 PAFDFKPFPAPGYAEDPFPAIERLREATPIF---YWDEGRSWVLTRYHDVSAVFRDERFAV 69
              |:|  ||       :|      |||   |:     :|   :|::|:| ||: |   |||
Db          6 PSF---PFQRASGMEPPAEFARLRATDPVSKVKLFDGSLAWLVTKYKDVTFVATDERL-- 60

Qy         70 SREEWESSAEYSSAIPELSDMKKYGLFGLP------PEDHARVRKLVNPSFTSRAIDLLR 123
                     |         |||:    |        |              ||    | :|    ||     :    |:
```

```
Db         61 ------SKVRTRPGFPELNAGGKQAAKAKPTFVDMDAPDHMNQRGMVESLFTLEHVKKLQ 114

Qy        124 AEIQRTVDQLLDARS----GQEEFDVVRDYAEGIPMRAISALLKVP------AECDEKFR 173
              ||:||| || |          |:|:::|  :|   |  :| ||           |
Db        115 PYIQKTVDDLLAAMKKKGCANGPVDLVKEFALPVPSYIIYTILGVPFNDLDHLTNQNAIR 174

Qy        174 RFGSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAE 233
              ||:|||                       |:  | |  |  ::::|  |  ::|::: |   :
Db        175 TNGSSTAREAS---------------AANQELLDYLASLVEKRLEEP-KDDLISKLCTEQ 218

Qy        234 ADGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDE 293
                 :  : |  :  :: ||  | : :||  |: | : ||  | :|| |   |     ::|
Db        219 VKPGNIEKADAVQIAFLLLVAGNATMVNMIALGVVTLFQHPEQLAQLKANPSLAPQFVEE 278

Qy        294 VLRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTG 353
              : |:      :      | |: |||  |  ||  | :          || ||  :|  || |:: |
Db        279 LCRYHTASALAIKRTAKVDLEIGGKHIKANEGIIASNQSANRDADIFENPDEFNMNRKWP 338

Qy        354 AS--LAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEM-------KLKETPVFG 398
              |   | || ||| |     ||: |       |:|: ||::         |:  ||: |
Db        339 AEDPLGYGFGPHRCIAEHLAKAELTTVFATLFKEFPDLNIAVPFEKINFTPLGG 392


RESULT 15
C128_MYCTU
ID   C128_MYCTU     STANDARD;      PRT;   489 AA.
AC   Q59572;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative cytochrome P450 128 (EC 1.14.-.-).
GN   CYP128 OR RV2268C OR MT2330 OR MTCY339.42.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z77163; CAB00967.1; -.
DR   EMBL; AE007076; AAK46612.1; -.
DR   PIR; H70729; H70729.
DR   HSSP; Q00441; 1OXA.
DR   TIGR; MT2330; -.
DR   TubercuList; Rv2268c; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW   Transmembrane; Complete proteome.
FT   TRANSMEM    207    227       POTENTIAL.
FT   TRANSMEM    256    276       POTENTIAL.
FT   TRANSMEM    315    335       POTENTIAL.
FT   TRANSMEM    383    403       POTENTIAL.
FT   METAL       435    435       IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ   SEQUENCE   489 AA;  53313 MW;  633F233CEF03AD7A CRC64;

  Query Match             17.6%;  Score 373.5;  DB 1;  Length 489;
  Best Local Similarity   28.3%;  Pred. No. 2.5e-16;
  Matches  115;  Conservative   65;  Mismatches  204;  Indels   23;  Gaps    6;

Qy         16 DFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWE 75
              || ||  |   | ||:|      |         :  |   :   ::|:|| ||      |:     |
Db        101 DFDPFDPAIAADPYPHYRELLAGERVQYNPKRDVYILSRYADVREAARNHDTLSSARGVT 160

Qy         76 SSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLD 135
                |  :   :|  ||             |  | |:|| : |       |::  |  :   :  :|:
Db        161 FSRGWLPFLP-TSD----------PPAHTRMRKQLAPGMARGALETWRPMVDQLARELVG 209

Qy        136 ARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEET 195
                 |   |||   |  :||||:::| |      :  |  |   :    |    |  |        :
Db        210 GLLTQTPADVVSTVAAPMPMRAITSVLGVDGPDEAAFCRLSNQAVRITDVAL-------S 262

Qy        196 KTLVASVTEGLA---LLHDVLDERRRNPLENDVLTM-LLQAEADGSRLSTKELVALVGAI 251
                : : |: :| |      |  :    || |  |   :     :   |    |:  |||  ||       :
Db        263 ASGLISLVQGFAGFRRLRALFTHRRDNGLLRECTVLGKLATHAEQGRLSDDELFFFAVLL 322

Qy        252 IAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQ 311
              : || ::| ::|:   | |    |:  |  |:  :|  :  :|::| ||| : ::     |   |
Db        323 LVAGYESTAHMISTLFLTLADYPDQLTLLAQQPDLIPSAIEEHLRFISPIQ-NICRTTRV 381

Qy        312 DLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLA 371
                  |    |  |  | :| |    :| ||    :  ||||    |:          ||:|  |  |:|||  ||
Db        382 DYSVGQAVIPAGSLVLLAWGAANRDPRQYEDPDVFRADRNPVGHLAFGSGIHLCPGTQLA 441

Qy        372 RLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKA 418
              |:|  :   :   |        ::: |  | :   :   |  :  |  |  :  |   |
Db        442 RMEGQAILREIVANIDRIEVVEPPTWTTNANLRGLTRLRVAVTPRVA 488

Search completed: October  2, 2003, 17:28:42
Job time : 9.74239 secs
```

```
OM protein - protein search, using sw model

Run on:          October  2, 2003, 17:27:33 ; Search time 14.5739 Seconds
                                              (without alignments)
                                              2764.850 Million cell updates/sec

Title:           US-10-014-717-8
Perfect score:   2124
Sequence:        1 MTQEQANQSETKPAFDFKPF..........HPAFRNIESLNVILKPSKAG 419

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:      283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_76:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	531	25.0	405	1	B42606	cytochrome P450 CV
2	525.5	24.7	395	1	G69594	cytochrome P450 bi
3	518.5	24.4	396	1	B69851	cytochrome P450 yj
4	502	23.6	410	1	E69611	cytochrome P450 cy
5	486	22.9	376	1	G69679	polyketide hydroxy
6	481.5	22.7	418	2	G83229	cytochrome P450 PA
7	459	21.6	403	2	B35401	cytochrome P450 10
8	451	21.2	415	2	AC2731	cytochrome P450 hy
9	451	21.2	415	2	B97512	probable cytochrom
10	450	21.2	402	1	A70707	cytochrome P450 Rv
11	442.5	20.8	434	2	C87170	probable cytochrom
12	434.5	20.5	438	1	E70515	cytochrome P450 Rv
13	430	20.2	411	1	JC4003	cytochrome P450 -
14	430	20.2	417	1	S49051	cytochrome P450 ty
15	429	20.2	411	2	T36526	probable cytochrom
16	421.5	19.8	444	2	C83336	probable cytochrom
17	420	19.8	411	1	JC5859	polyketide synthas
18	415.5	19.6	402	2	H82813	cytochrome P450-li
19	413	19.4	397	1	S51594	cytochrome P450 my
20	411	19.4	404	2	T30231	cytochrome P450 -
21	406	19.1	310	2	T44767	cytochrome p450 [i
22	400	18.8	406	1	S18531	cytochrome P450 er
23	386.5	18.2	408	2	PD0007	cytochrome P450 no
24	386.5	18.2	408	2	JC5674	cytochrome P450 no
25	386	18.2	398	1	H70807	cytochrome P450 Rv
26	380	17.9	381	1	S15809	cytochrome P450 CY
27	373.5	17.6	489	1	H70729	cytochrome P450 Rv
28	366	17.2	410	2	S39924	cytochrome P450meg
29	364.5	17.2	399	1	A55578	cytochrome P450 -
30	361.5	17.0	412	2	S24750	cytochrome P450 (s
31	361.5	17.0	433	2	T44587	cytochrome P450 ho
32	357.5	16.8	410	2	JC4287	cytochrome P450 sc
33	356.5	16.8	400	2	E82817	cytochrome P-450 h
34	354.5	16.7	406	2	A35401	cytochrome P450 10
35	353.5	16.6	402	2	JC5151	nitric-oxide reduc
36	353.5	16.6	404	1	JC5150	nitric-oxide reduc
37	352.5	16.6	405	1	F69611	cytochrome P450 cy
38	350	16.5	401	1	I40208	cytochrome P450 BJ
39	344.5	16.2	400	1	H70921	cytochrome P450 Rv
40	344.5	16.2	433	1	B70677	cytochrome P450 Rv
41	342.5	16.1	428	1	A42971	cytochrome P450ter
42	338.5	15.9	444	2	F87366	cytochrome P450 fa
43	332	15.6	410	1	O4BS6M	cytochrome P450 10
44	328.5	15.5	406	1	A48495	linalool 8-monooxy
45	328.5	15.5	467	2	T10879	y41C protein - Rhi

ALIGNMENTS

```
RESULT 1
B42606
cytochrome P450 CVIIB1 - Saccharopolyspora erythraea
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Saccharopolyspora erythraea
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C;Accession: B42606
R;Andersen, J.F.; Hutchinson, C.R.
J. Bacteriol. 174, 725-735, 1992
A;Title: Characterization of Saccharopolyspora erythraea cytochrome P-450 genes and enzy
A;Reference number: A42606; MUID:92121109; PMID:1732208
A;Accession: B42606
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA; protein
A;Residues: 1-405 <AND>
A;Cross-references: GB:M83110; NID:g152682; PIDN:AAA26483.1; PID:g152684
A;Experimental source: NRRL2338
A;Note: sequence extracted from NCBI backbone (NCBIP:77484)
C;Genetics:
A;Gene: CYP107B1
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F;238-374/Domain: cytochrome P450 homology <CYP>
F;352/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match             25.0%;  Score 531;  DB 1;  Length 405;
  Best Local Similarity   32.5%;  Pred. No. 6.8e-30;
  Matches  132;  Conservative   83;  Mismatches  169;  Indels   22;  Gaps    6;

Qy         16 DFKPFAPGYAEDPFPAIERLRE--ATPIFYWDEGRSWVLTRYHDVSAVFRDER----FAV 69
               |   |   :|:|     |:||     |    :    :|::||| ||     | |   |
Db          8 DLLAFDDAFAQDRHNRYARMREEPVQRIRTVNGLDAWLITRYEDVKQALLDPRIAKDFGR 67

Qy         70 SREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRT 129
              ::: |    :   |  |     :      | || |:||||  :||:| :: ||  |::
Db         68 TQQIIEKRLADAERRPGFSPDLGPHMLNTDPPDHTRLRKLVVKAFTARRVEGLRPRIEQI 127

Qy        130 VDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVP 189
               | |||  :|: | |:: ::|  :|:  || |: |         : || :        |  |  |
Db        128 TDDLLDRLAGRSEVDLIDEFAFPLPITVISELMGVEDSRRDDFRSW----TNVLVDGSQP 183

Qy        190 QVDEETK-TLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALV 248
                :     :|  :|| :|           ::|     :|:|| ||:|  || |||  ||:|:|
Db        184 EAQAQASVAMVEYLTELIA--------KKRTEPGDDLLTALLEAVEDGDRLSEGELIAMV 235

Qy        249 GAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRF 308
                :: || :||: ||    ||:|| :|:  |    :: :| |:  |::| ||::: :  || |
Db        236 FLLLVAGHETTVNLIGNCVLSLLGNPDQLAALRNDPSLLPGAIEETLRYESPVANGTFRH 295

Qy        309 ARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGV 368
                : : :      | :||:|:: : :| |||   |   || ||: |:|     :|:| | | | |
Db        296 TAEAVRFGDVVIPEGELVWVALGAANRDGERFEDPDRFDITRETTGHVAFGHGIHFCVGA 355
```

```
Qy        369 SLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
              :||||||:||||  :   |||::::   :|      : :      |:| | |
Db        356 ALARLEAQIAVGRLLERFPDLRMAASPDDLRWRFSVLMRGLEKLPV 401
```

RESULT 2
G69594
cytochrome P450 bioI - Bacillus subtilis
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Bacillus subtilis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: G69594
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chc
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis,
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: G69594
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-395 <KUN>
A;Cross-references: GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CAB14997.1; PID:g2635503
A;Experimental source: strain 168
C;Genetics:
A;Gene: bioI
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: oxidoreductase
F;232-367/Domain: cytochrome P450 homology <CYP>

```
  Query Match             24.7%;  Score 525.5;  DB 1;  Length 395;
  Best Local Similarity   31.9%;  Pred. No. 1.6e-29;
  Matches  129;  Conservative   98;  Mismatches  137;  Indels   41;  Gaps   12;

Qy         24 YAEDPFPAIERLREATPIFYWDEGR-----SWVLTRYHDVSAVFRDERFAVSREEWESSA 78
               : ::|:    : ||   ||:  :         | :|  | : :|: :| || |     |||
Db         12 FLKNPYSFYDTLRAVHPIY---KGSFLKYPGWYVTGYEETAAILKDARFKVRTPLPESST 68

Qy         79 EYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARS 138
              :|    :|| ::  :      || |:| :||  :     :     :  ||  |||
Db         69 KYQ----DLSHVQNQMMLFQNQPDHRRLRTLASGAFTPRTTESYQPYIIETVHHLLDQVQ 124

Qy        139 GQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTL 198
              |:::  :|: |:|   :      |: :: ||  |   |: : : ::          |:  :|  |::
Db        125 GKKKMEVISDFAFPLASFVIANIIGVPEEDREQLKEWAAS--------LIQTID-FTRSR 175

Qy        199 VASVTEG-------LALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAI 251
                | :|||          :|   ::: :|:|:| : |:::|||:   :   :|: :|   :    :
Db        176 KA-LTEGNIMAVQAMAYFKELIQKRKRHP-QQDMISMLLKGR-EKDKLTEEEAASTCILL 232

Qy        252 IAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQ 311
                 || :||: ||: :|| ||: || |    ::   | |:   |::| ||:::  :: | | | :
Db        233 AIAGHETTVNLISNSVLCLLQHPEQLLKLRENPDLIGTAVEECLRYESPTQM-TARVASE 291

Qy        312 DLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLA 371
              |:: || :|::|| |:||: :| || ::|: |||||:  |     |::| | ||| | |||
Db        292 DIDICGVTIRQGEQVYLLLGAANRDPSIFTNPDVFDITRSPNPHLSFGHGHHVCLGSSLA 351

Qy        372 RLEAEIAVGTIFRRFPEMKLKE-----TPVFGYHPAFRNIESLNV 411
              ||||:||: |: :| | : |  :          |:||       || :| | |
Db        352 RLEAQIAINTLLQRMPSLNLADFEWRYRPLFG----FRALEELPV 392
```

RESULT 3
B69851
cytochrome P450 yjiB - Bacillus subtilis
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Bacillus subtilis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: B69851
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chc
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B69851
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-396 <KUN>
A;Cross-references: GB:Z99110; GB:AL009126; NID:g2633472; PIDN:CAB13078.1; PID:g2633575
A;Experimental source: strain 168
C;Genetics:
A;Gene: yjiB
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: oxidoreductase
F;236-371/Domain: cytochrome P450 homology <CYP>

```
  Query Match             24.4%;  Score 518.5;  DB 1;  Length 396;
  Best Local Similarity   32.6%;  Pred. No. 5e-29;
  Matches  124;  Conservative   72;  Mismatches  155;  Indels   29;  Gaps    6;

Qy         25 AEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAI 84
               | |||  | :|:  |: : :| : | :   | ||  |  |:               :|| :
Db         23 AYHPFPWYESMRKDAPVSFDEENQVWSVFLYDDVKKVVGDKEL-----------FSSCM 70

Qy         85 PELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEEFD 144
              |: :        :  : |   | ::| :|| :|| | :        ||     ||:|:    |: |||
Db         71 PQQTSSIGNSIINMDPPKHTKIRSVVNKAFTPRVMKQWEPRIQEITDELIQKFQGRSEFD 130

Qy        145 VVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLVA---- 200
              :| |:  :|:  || || ||:    |:|: :                    ||:  :
Db        131 LVHDFSYPLPVIVISELLGVPSAHMEQFKAWSD-----LLVSTPKDKSEEAEKAFLEERD 185

Qy        201 SVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTI 260
                  | ||        : | :|| | |:::|::|| | :|| :||:          :: || :||
Db        186 KCEEELAAFFAGIIEEKRNKPEQDIISILVEAEETGEKLSGEELIPFCTLLLVAGNETTT 245

Qy        261 YLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRF---DNILRIGTVRFARQDLEYCG 317
                ||: |: ::| :|    | ::: | || |::| |||      :||    | |::| | |
Db        246 NLISNAMYSILETPGVYEELRSHPELMPQAVEEALRFRAPAPVLR----RIAKRDTEIGG 301

Qy        318 ASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLEAEI 377
                ||:|:||   : || ||  | || :||:||       :|:| | | | | |  ||||| |
Db        302 HLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHPNPHIAFGHGIHFCLGAPLARLEANI 361

Qy        378 AVGTIFRRFPEMK-LKETPV 396
              |: ::   || |: :   ||:
Db        362 ALTSLISAFPHMECVSITPI 381
```

RESULT 4
E69611
cytochrome P450 cypA - Bacillus subtilis
N;Contains: oxidoreductase (EC 1.-.-.-)

```
C;Species: Bacillus subtilis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: E69611; T44774
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chc
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69611
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-410 <KUN>
A;Cross-references: GB:Z99117; GB:AL009126; NID:g2634966; PIDN:CAB14615.1; PID:g2635119
A;Experimental source: strain 168
R;Belitsky, B.R.; Gustafsson, M.C.U.; Sonenshein, A.L.; von Wachenfeldt, C.
J. Bacteriol. 179, 5448-5457, 1997
A;Title: An lrp-like gene of Bacillus subtilis involved in branched-chain amino acid tra
A;Reference number: Z22837; MUID:97431495; PMID:9287000
A;Accession: T44774
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-410 <BEL>
A;Cross-references: EMBL:Y11043; NID:g1926275; PIDN:CAA71937.1; PID:g1926278
A;Experimental source: strain 1A1
C;Genetics:
A;Gene: cypA
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: oxidoreductase
F;245-381/Domain: cytochrome P450 homology <CYP>

  Query Match             23.6%;  Score 502;  DB 1;  Length 410;
  Best Local Similarity   30.9%;  Pred. No. 7.7e-28;
  Matches  115;  Conservative   88;  Mismatches  159;  Indels   10;  Gaps    4;

Qy         24 YAEDPFPAIERLREAT---PIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAEY 80
              : :: :   : ||::    |:      |: |:::|| |    : ::|:    : |    :: |
Db         22 FKDEAYEFYKELRKSQALYPLSLGALGKGWLISRYDDAIHLLKNEKLKKNYENVFTAKEK 81

Qy         81 SSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQ 140
                |: :  :      :      | || |:| ||  :|| | |  |   :||   | |||
Db         82 RPALLKNEETLTKHMLNSDPPDHNRLRTLVQKAFTHRMILQLEDKIQHIADSLLDKVQPN 141

Qy        141 EEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLVA 200
              :  ::| |||  :|:  || :| :| |  :||| :  |     :       |:  :|   |:
Db        142 KFMNLVDDYAFPLPIIVISEMLGIPLEDRQKFRVWSQA---IIDFSDAPERLQENDHLLG 198

Qy        201 SVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTI 260
                 |    |  ::  ::|| |    |:::  |:|||::|::|||:||  :::   :|  || :||:
Db        199 EFVE---YLESLVRKKRREP-AGDLISALIQAESEGTQLSTEELYSMIMLLIVAGHETTV 254

Qy        261 YLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEYCGASI 320
                ||      |:    :  ||  :: :|  ||  :|::| |||  :  :  :  |:|:   :      |  |
Db        255 NLITNMTYALMCHHDQLEKLRQQPDLMNSAIEEALRFHSPVELTTIRWTAEPFILHGQEI 314

Qy        321 KKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLEAEIAVG 380
              |: ::: :  :  || ||   :|    |:||:  |        :|:|  |  |  |  |   |||||||:||:
Db        315 KRKDVIIISLASANRDEKIFPNADIFDIERKNNRHIAFGHGNHFCLGAQLARLEAKIAIS 374

Qy        381 TIFRRFPEMKLK 392
              |:  || |  ::||
Db        375 TLLRRCPNIQLK 386
```

```
RESULT 5
G69679
polyketide hydroxylase (EC 1.-.-.-) pksS - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: G69679
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chc
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: G69679
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-376 <KUN>
A;Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CAB13607.1; PID:g2634107
A;Experimental source: strain 168
C;Genetics:
A;Gene: pksS
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: oxidoreductase
F;240-376/Domain: cytochrome P450 homology <CYP>

  Query Match             22.9%;  Score 486;  DB 1;  Length 376;
  Best Local Similarity   31.6%;  Pred. No. 9.2e-27;
  Matches  120;  Conservative   76;  Mismatches  156;  Indels   28;  Gaps    7;

Qy         17 FKPFAPGYAEDPFPAIERLREATPIFYWDEGR------SWVLTRYHDVSAVFRDERFA-- 68
              | |     :  :||  : | ||  ||  ::  |      :|::|||  |   |  :| |
Db          8 FHPHGKEFHHNPFSVLGRFREEEPIHRFELKRFGATYPAWLITRYDDCMAFLKDNRITRD 67

Qy         69 ---VSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAE 125
                 |  :|        |  |  :||       :          || |:| ||: :|| |  |: ||
Db         68 VKNVMNQEQIKMLNVSEDIDFVSD----HMLAKDTPDHTRLRSLVHQAFTPRTIENLRGS 123

Qy        126 IQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARAL-- 183
              |::  :||||    : : |::: :|   :|    || |: :| |    ::|: : :|
Db        124 IEQIAEQLLDEMEKENKADIMKSFASPLPFIVISELMGIPKEDRSQFQIWTNAMVDTSEG 183

Qy        184 GVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKE 243
                 |  |   | |  :|         |:||     ||  | ::|:::  |: || :||:|| ||
Db        184 NRELTNQALREFKDYIAK------LIHD----RRIKP-KDDLISKLVHAEENGSKLSEKE 232

Qy        244 LVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRI 303
              | :::   ::  || :||: |:      | ||:   :   | :| :|  ::   |::|:||: : : :
Db        233 LYSMLFLLVVAGLETTVNLLGSGTLALLQHKKECEKLKQQPEMIATAVEELLRYTSPVVM 292

Qy        304 GTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPH 363
                 |:| :|    |  |  |||:|:|:|:  | || ||    |  |::  ::  |       :::| | |
Db        293 MANRWAIEDFTYKGHSIKRGDMIFIGIGSANRDPNFFENPEILNINRSPNRHISFGFGIH 352

Qy        364 VCPGVSLARLEAEIAVGTIF 383
               | |  |||||  ||     |
Db        353 FCLGAPLARLEGHIAFKAAF 372

RESULT 6
G83229
cytochrome P450 PA3331 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 16-Feb-2001
```

```
C;Accession: G83229
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathc
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: G83229
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-418 <STO>
A;Cross-references: GB:AE004755; GB:AE004091; NID:g9949456; PIDN:AAG06719.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3331
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
F;233-369/Domain: cytochrome P450 homology <P45>

  Query Match           22.7%;  Score 481.5;  DB 2;  Length 418;
  Best Local Similarity  33.5%;  Pred. No. 2.2e-26;
  Matches  116;  Conservative   73;  Mismatches  140;  Indels   17;  Gaps    5;

Qy         50 WVLTRYHDVSAVFRDERFAVSREEWESSAEYS----SAIPELSDMKKYGLFGLPPEDHAR 105
Db         41 WVVTRYRDARKVLNHP--GVRRDARQAAELYAKRTGSPRAGIGEGLSHHMLNLDPPDHTR 98

Qy        106 VRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVP 165
Db         99 LRSLVGRAFTPRQVERLQPHIERITEALLDAMAGREQADLMADFAIPLTIAVIFELLGIP 158

Qy        166 AECDEKFRRFGSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDV 225
Db        159 EAEREHARQSWERQAELLS-------PEEAQALADAQVDYLRVL---LEAKRRQPAD-DV 207

Qy        226 LTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPG 285
Db        208 YSGLVQAADESGQLSEAELVSMAHLLMMSGFETTMNMIGNALVTLLVNPEQLALLRAQPE 267

Qy        286 LMRNALDEVLRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDV 345
Db        268 LLPNAMEELVRHDSPVRASMLRFTVEDVELDGVTIPAGEYILVSNLTANHDAERFDDPDR 327

Qy        346 FDVRRDTGASLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKL 391
Db        328 LDLTRNTDGHLGYGFGVHYCVGASLARLEGRIAIQRLLARFPDLQL 373
```

```
RESULT 7
B35401
cytochrome P450 105B1 - Streptomyces griseolus
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Streptomyces griseolus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 04-Mar-2000
C;Accession: B35401
R;Omer, C.A.; Lenstra, R.; Litle, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Romesser,
J. Bacteriol. 172, 3335-3345, 1990
A;Title: Genes for two herbicide-inducible cytochromes P-450 from Streptomyces griseolus
A;Reference number: A35401; MUID:90264332; PMID:2345149
A;Accession: B35401
A;Molecule type: DNA
A;Residues: 1-403 <OME>
A;Cross-references: GB:M32239; NID:g153480; PIDN:AAA26825.1; PID:g153481; GB:M36481
C;Genetics:
A;Gene: CYP105B1
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;
F;238-374/Domain: cytochrome P450 homology <P45>
F;352/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match           21.6%;  Score 459;  DB 2;  Length 403;
  Best Local Similarity  31.7%;  Pred. No. 8.1e-25;
  Matches  127;  Conservative   73;  Mismatches  157;  Indels   44;  Gaps   12;

Qy         21 APGYAEDPFPAIERLREATPI---FYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESS 77
Db         20 APGCPFDPAPDVTEAARTEPVTRATLWDGSSCWLVTRHQDVRAVLGDPRF--------SA 71

Qy         78 AEYSSAIPELSDMKKYGLFGLPP-------EDHARVRKLVNPSFTSRAIDLLRAEIQRTV 130
Db         72 DAHRTGFPFLTAGGR-EIIGTNPTFLRMDDPEHARLRRMLTADFIVKKVEAMRPEVQRLA 130

Qy        131 DQLLD-ARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVP 189
Db        131 DDLVDRMTTGRTSADLVTEFALPLPSLVICLLLGVPYE-DHAF--FQERSRVLLTLRSTP 187

Qy        190 QVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVG 249
Db        188 ---EEVRAAQDELLEYLARL---ARTKRERPDDAIISRLVARGELDDTQIATMGRLLLV- 240

Qy        250 AIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFA 309
Db        241 ----AGHETTANMTALSTLVLLRNPDQLARLRAEPALVKGAVEELLRYLTIVHNGVPRIA 296

Qy        310 RQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVS 369
Db        297 TEDVLIGGRTIAAGEGVLCMISSANRDAEVFPGGDDLDVARDARRHVAFGFGVHQCLGQP 356

Qy        370 LARLEAEIAVGTIFRRFPEMKL----KETP------VFGYH 400
Db        357 LARVELQIAIETLLRRLPDLRLAVPHEEIPFRGDMAIYGVH 397
```

```
RESULT 8
AC2731
cytochrome P450 hydroxylase Atu1256 [imported] - Agrobacterium tumefaciens (strain C58,
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AC2731
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AC2731
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-415 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL42265.1; PID:g17739662; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu1256
A;Map position: circular chromosome
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: heme; iron; metalloprotein
F;367/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match           21.2%;  Score 451;  DB 2;  Length 415;
  Best Local Similarity  29.1%;  Pred. No. 3.1e-24;
  Matches  118;  Conservative   82;  Mismatches  175;  Indels   30;  Gaps    9;

Qy         22 PGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERF------AVSREEWE 75
Db         24 PAFYNDPNPVYAALHAQCPTFYWEEQRQWFFTCYDHVSTLLRDRRFGRQILHVASREE-- 81

Qy         76 SSAEYSSAIPE-LSDMK------KYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQR 128
Db         82 ------IGLPEPLEHVKHFDLAEQHSLLELEPPEHTRLRTLINRAFVSRHVDKMKPEIEE 135

Qy        129 TVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLV 188
Db        136 LANRLIEAFEANGETELLSSYADIIPVTMIARMIGIPEEMGPQLLKWSHA---YVGMYMF 192

Qy        189 PQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALV 248
```

```
              :  |:         :   |      :   |: |||  | ::|:|: ::    |    |    |:   |||:
Db        193 KRTPEDELLADKAAQEFSDYVRRVIAERRAEP-KDDLLSHMIHTEHKGQYLTDDELVSTT 251

Qy        249 GAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRF 308
              :: || : |:: |   :|   :|  |        :  :   :       ::|  ||       :  |    |:
Db        252 IVLLNAGHEATVHQIGNSVRIILESGLDPKTLFHDETATERTVEETLRICAPVHI-FQRW 310

Qy        309 ARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGV 368
               : :|   |    |:|: | |:: :|    |    || |    |     |: ||::::| | | | |
Db        311 VLEPVEIDGVQFKRGDKVSLILAAANLDPAKFSDPLAFQPDRNEGANVSFGAGIHFCIGA 370

Qy        369 SLARLEAEIAVGTIFRRFPEMKLKETPVFG--YHPAFRNIESLNV 411
              |||||  :|:   :|:|  | :|: | |        ||   |   :| |::
Db        371 PLARLELNLALPLLFKRLPGLKIAEPPKVKDVYH--FHGLERLDL 413
```

RESULT 9
B97512
probable cytochrome P450 hydroxylase [imported] - Agrobacterium tumefaciens (strain C58,
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: B97512
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: B97512
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-415 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87051.1; PID:g15156303; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_2319
A;Map position: circular chromosome
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: heme; iron; metalloprotein
F;367/Binding site: heme iron (Cys) (axial ligand) #status predicted

```
  Query Match             21.2%;  Score 451;  DB 2;  Length 415;
  Best Local Similarity   29.1%;  Pred. No. 3.1e-24;
  Matches  118;  Conservative   82;  Mismatches  175;  Indels   30;  Gaps    9;

Qy         22 PGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERF------AVSREEWE 75
              | :   ||        |     ||||:| |   ||    || : |||              ||||
Db         24 PAFYNDPNPVYAALHAQCPTFYWEEQRQWFFTCYDHVSTLLRDRRFGRQILHVASREE-- 81

Qy         76 SSAEYSSAIPE-LSDMK------KYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQR 128
                    :|| |   :|            :: |    | | :| |:| |:| :| || :| :: ||:
Db         82 ------IGLPEPLEHVKHFDLAEQHSLLELEPPEHTRLRTLINRAFVSRHVDKMKPEIEE 135

Qy        129 TVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLV 188
              ::|::|       | :::  ||: ||:  |: :: :| |     :   ::   |       :|: :
Db        136 LANRLIEAFEANGETELLSSYADIIPVTMIARMIGIPEEMGPQLLKWSHA---YVGMYMF 192

Qy        189 PQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALV 248
              :  |:        :      :   |: |||   | ::|:|: ::    |    |    |:   |||:
Db        193 KRTPEDELLADKAAQEFSDYVRRVIAERRAEP-KDDLLSHMIHTEHKGQYLTDDELVSTT 251

Qy        249 GAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRF 308
              :: || : |:: |   :|   :|  |        :  :   :       ::|  ||       :  |    |:
Db        252 IVLLNAGHEATVHQIGNSVRIILESGLDPKTLFHDETATERTVEETLRICAPVHI-FQRW 310

Qy        309 ARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGV 368
               : :|   |    |:|: | |:: :|    |    || |    |     |: ||::::| | | | |
Db        311 VLEPVEIDGVQFKRGDKVSLILAAANLDPAKFSDPLAFQPDRNEGANVSFGAGIHFCIGA 370

Qy        369 SLARLEAEIAVGTIFRRFPEMKLKETPVFG--YHPAFRNIESLNV 411
              |||||  :|:   :|:|  | :|: | |        ||   |   :| |::
Db        371 PLARLELNLALPLLFKRLPGLKIAEPPKVKDVYH--FHGLERLDL 413
```

RESULT 10

A70707
cytochrome P450 Rv0766c - Mycobacterium tuberculosis (strain H37RV)
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Mycobacterium tuberculosis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: A70707
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70707
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-402 <COL>
A;Cross-references: GB:Z80226; GB:AL123456; NID:g3261638; PIDN:CAB02396.1; PID:g1550644
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0766c
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: oxidoreductase
F;236-372/Domain: cytochrome P450 homology <CYP>

```
  Query Match             21.2%;  Score 450;  DB 1;  Length 402;
  Best Local Similarity   32.4%;  Pred. No. 3.5e-24;
  Matches  128;  Conservative   64;  Mismatches  179;  Indels   24;  Gaps    9;

Qy         13 PAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRD-----ERF 67
              |     |:     : |||:|    |||:   |::  :|     | ::|:|||   |||          :
Db          8 PELVLDPYDYDFHEDPYPYYRRLRDEAPLYRNEERNFWAVSRHHDVLQGFRDSTALSNAY 67

Qy         68 AVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQ 127
              ||       : |:   |  |    :|          :    :       | |:| ||:   ||  |  |   ::
Db         68 GVS---LDPSSRTSEAYRVMS------MLAMDDPAHLRMRTLVSKGFTPRRIRELEPQVL 118

Qy        128 RTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA-RALGVG 186
                        ||:     |  || | ::|  :||  || |: ||          : | .    |     ||
Db        119 ELARIHLDSALQTESFDFVAEFAGKLPMDVISELIGVPDTDRARIRALADAVLHREDGVA 178

Qy        187 LVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVA 246
                  ||          : ||: | :       |::  | || |     |:: : || || || ||| :|::|
Db        179 DVP-----PPAMAASI-ELMRYYADLIAEFRRRP-ANNLTSALLAAELDGDRLSDQEIMA 231

Qy        247 LVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTV 306
              :  :: || :||   |:| ||             |    |    | |:    :    ::| ||:|     :|
Db        232 FLFLMVIAGNETTTKLLANAVYWAAHHPGQLARVFADHSRIPMWVEETLRYDTSSQI-LA 290

Qy        307 RFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASL-AYGRGPHVC 365
                    ||              :| :||:: ||   || ||  ||.|| : : |:  |   ::| | |
Db        291 RTVAHDLTLYDTTIPEGEVLLLLPGSANRDDRVFDDPDDYRIGREIGCKLVSFGSGAHFC 350

Qy        366 PGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYH 400
                |  |||:|| :|:|  :  ||      :: :   |    |
Db        351 LGAHLARMEARVALGALLRRIRNYEVDDDNVVRVH 385
```

RESULT 11
C87170
probable cytochrome p450 (imported) - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: C87170
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: C87170
A;Status: preliminary

```
A;Molecule type: DNA
A;Residues: 1-434 <STO>
A;Cross-references: GB:AL450380; NID:g13093682; PIDN:CAC31043.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML2088
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

  Query Match             20.8%;  Score 442.5;  DB 2;  Length 434;
  Best Local Similarity   31.4%;  Pred. No. 1.3e-23;
  Matches  126;  Conservative   73;  Mismatches  179;  Indels   23;  Gaps    9;

Qy         22 PGYAEDPFPAIERLREATPIFYWDEGRS-WVLTRYHDVSAVFRDERFAVSREEWESSAEY 80
Db         44 PGTRADPFPVYRALIDYGPMQL--PGMPLTVFSSFSDCDEALRHPLSASDRLKATLAQQA 101

Qy         81 SSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQ 140
Db        102 IAAGAEPRPFYASSFMFLDPPDHTRLRKLVSKAFAPKVVQALEGDIAALVDSLLDKGAAA 161

Qy        141 EEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALG-----VGLVPQVDEET 195
Db        162 GQFDVIADLAFPLAVAVICRLLGVPYEDAPEFGRVSALLVQSVDPFITITGEPPEATEE- 220

Qy        196 KTLVASVTEGLALLHDVLDE----RRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAI 251
Db        221 -RLRAGV-----WLRDYLEQLVKCRRGTPGE-DLISRLIELDESGDQLTEEEIIATCGLL 273

Qy        252 IAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILR-IGTVRFAR 310
Db        274 LVAGHETTVNLIANAVLAMLRNPSQWKALSSNPQRAPLVVEETLRYDPAIHLIG--RVAA 331

Qy        311 QDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSL 370
Db        332 KDMTIGQTTLTEGDTMVLLLAAANRDPAVYSRPDEFDPDRPSSRHLAFAVGSHFCLGAAL 391

Qy        371 ARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNV 411
Db        392 ARLEATVTLSAISARFPQVQLAGELVYKPNVAMRGMSALPV 432

RESULT 12
E70515
cytochrome P450 Rv1880c - Mycobacterium tuberculosis (strain H37RV)
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Mycobacterium tuberculosis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: E70515
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: E70515
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-438 <COL>
A;Cross-references: GB:Z97193; GB:AL123456; NID:g3261816; PIDN:CAB10066.1; PID:g2225980
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1880c
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F;267-403/Domain: cytochrome P450 homology <P45>
F;381/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match             20.5%;  Score 434.5;  DB 1;  Length 438;
  Best Local Similarity   29.6%;  Pred. No. 4.8e-23;
  Matches  126;  Conservative   77;  Mismatches  168;  Indels   55;  Gaps   13;

Qy         22 PGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVS-AVFRDERFAVSREEWESSAEY 80
Db         35 PAVATDPVPFYDEVRSHGALV---RNRANYLTVDHRLAHDLLRSDDFRV--------VSF 83

Qy         81 SSAIP------------ELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQ 127
Db         84 GENLPPPLRWLERRTRGDQLHPLREPSLLAVEPPDHTRYRKTVSAVFTSRAVSALRDLVE 143

Qy        128 RTVDQLLDARSGQEEF-DVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVG 186
Db        144 QTAINLLDRFAEQPGIVDVVGRYCSQLPIVVISEILGVPEHDRPRVLEFGELAAPSLDIG 203

Qy        187 LVPQVDEETKTLVASVTEGL----ALLHDVLDERRRNPLENDVLTMLLQAEADG---SRL 239
Db        204 I-------PWRQYLRVQQGIRGFDCWLEGHLQQLRHAP-GDDLMSQLIQIAESGDNETQL 255

Qy        240 STKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDN 299
Db        256 DETELRAIAGLVLVAGFETTVNLLGNGIRMLLDTPEHLATLRQHPELWPNTVEEILRLDS 315

Qy        300 ILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRR-DTGASLAY 358
Db        316 PVQL-TARVACRDVEVAGVRIKRGEVVVIYLAAANRDPAVFPDPHRFDIERPNAGRHLAF 374

Qy        359 GRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKL------KETPVFGYHPAFRNIESLNVI 412
Db        375 STGRHFCLGAALARAEGEVGLRTFFDRFPDVRAAGAGSRRDTRV------LRGWSTLPVT 428

Qy        413 LKPSKA 418
Db        429 LGPARS 434

RESULT 13
JC4003
cytochrome P450 - Streptomyces sp.
N;Contains: oxidoreductase (EC 1.-.-.-)
C;Species: Streptomyces sp.
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: JC4003
R;Arisawa, A.; Tsunekawa, H.; Okamura, K.; Okamoto, R.
Biosci. Biotechnol. Biochem. 59, 582-588, 1995
A;Title: Nucleotide sequence analysis of the carbomycin biosynthetic genes including the
A;Reference number: JC4001; MUID:95290751; PMID:7772821
A;Accession: JC4003
A;Molecule type: DNA
A;Residues: 1-411 <ARI>
A;Cross-references: DDBJ:D30759; NID:g551628; PIDN:BAA06420.1; PID:g551630
A;Note: the source was designated as Streptomyces thermotolerans
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: oxidoreductase
F;242-382/Domain: cytochrome P450 homology <CYP>

  Query Match             20.2%;  Score 430;  DB 1;  Length 411;
  Best Local Similarity   29.6%;  Pred. No. 9.2e-23;
  Matches  120;  Conservative   75;  Mismatches  155;  Indels   56;  Gaps   12;

Qy         22 PGYAEDPFPAIERLREATPI----FYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESS 77
Db          9 PEVLRDPFGAYAEIRAQAPLVRASYPWG-AVQWLATRYSDVKAVLTDPRL---------- 57

Qy         78 AEYSSAIPELSDMKKY----GLFGLPPE--------------DHARVRKLVNPSFTSRA 118
Db         58 VNNPANVPEMHLPHPYEQALGDGGIPDEYVRYLAGSILSQDGPGHLRLRRLSSRAFTVRR 117

Qy        119 IDLLRAEIQRTVDQLLDA---RSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRF 175
Db        118 VNALRPRVTELAHRLLASLPDRAQDGVIDVLEDFSYPLSIDVICEIVGIPEEAREQWHTW 177

Qy        176 GSATARALGVGLVPQVDEETKTLVASVTEGLA-LLHDVLDERRRNPLENDVLTMLLQAEA 234
Db        178 GSAF-----YTMDPAV-------IGPAVRGMADHLHLLIEQRRATP-TGDLLTGLVQAED 224

Qy        235 D-GSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDE 293
```

```
Db        225 EQGEPLTDEEIATLVLTFVTAGNETTAHLIGNGVAALLTHSDQLALLRSDRRLLSQAVDE 284

Qy        294 VLRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRD-- 351
              ::|:   :::   |:| :||:  | :::||| |  :| :|  |    |   |: ||: |:
Db        285 LMRWCTPVQVTQPRYATEDLDVGGVTVRKGEQVVAVIGAAGHDPDRFPDPERFDITRNHR 344

Qy        352 --TGASLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP 395
                | : :| ||| | | :|| |  ||: |:| ||| : |   |
Db        345 APHEAHVGFGFGPHYCLGAALAHQETAIALDTLFDRFPSLALAVPP 390


RESULT 14
S49051
cytochrome P450 tylI (EC 1.1.-.-) [validated] - Streptomyces fradiae (strain T59235)
C;Species: Streptomyces fradiae
A;Variety: strain T59235
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: S49051
R;Merson-Davies, L.A.; Cundliffe, E.
Mol. Microbiol. 13, 349-355, 1994
A;Title: Analysis of five tylosin biosynthetic genes from the tyIIBA region of the Strep
A;Reference number: S49051; MUID:95075319; PMID:7984112
A;Accession: S49051
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-417 <MER>
A;Cross-references: EMBL:U08223; NID:g6849140; PIDN:AAA21341.1; PID:g473597
A;Experimental source: strain T59235
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
C;Genetics:
A;Gene: tylI
C;Function:
A;Description: involved in C20ring oxidation of O-mycaminosyl tylactone [validated, MUID
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: heme; oxidoreductase
F;253-388/Domain: cytochrome P450 homology <CYP>

  Query Match             20.2%;  Score 430;  DB 1;  Length 417;
  Best Local Similarity   32.0%;  Pred. No. 9.4e-23;
  Matches  120;  Conservative   70;  Mismatches  161;  Indels   24;  Gaps    8;

Qy         34 RLREATPIFYWDEGR-SWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAIP---ELSD 89
              | |    |:| ||| | :| : ||              |:    ||       |:::|    |    ||:
Db         41 RARAEAPVF-WDESRQAWQVFRY--------DDYLTVSTNPQLFSSDFSPVFPVPEELAI 91

Qy         90 MKKYGLF-GLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEEFDVVRD 148
              :    | | |: |   :|||: :|| | |  |     |     |||     : : ||| |
Db         92 LMGPGTFGGIDPPRHGPLRKLVSQAFTPRRIATLEPRIAEITRGLLDGLREKGQIDVVSD 151

Qy        149 YAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLVASVTEGLAL 208
              |  :|:  || :|||  :  || :                |:         |:  ::|:  :: |
Db        152 LAYPLPVIVIAELLGIPAEDRDLFREWVDVILNNEGMEYPNLPDDFSETMGPAIKEWGDY 211

Qy        209 LHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVL 268
              |:  :  :|  | : |::: |::|| :| ||: :|:| :|  ::  ||  ::  |:   |
Db        212 LYRRIALKRETPTD-DLMSGLIEAEVEGRRLTDEEIVNIVALLLTAGHISSATLLGNLFL 270

Qy        269 NLLRSPEALELVKAEPGLMRNALDEVLR----FDNILRIGTVRFARQDLEYCGASIKKGE 324
              |     ||   ::|:  |:  |::| ||     |:|| |:      ::| :  |  :| |:
Db        271 VLDEHREAQAELRADRDLIPGAIEETLRYRSPFNNIFRL-----LKEDTDILGHPMKAGQ 325

Qy        325 MVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLEAEIAVGTIFR 384
              ||    | || ||    || || |||||      :::| | | | |  |||||::  :    |
Db        326 MVVAWIASANRDSAHFSDPDTFDVRRQPNKHMSFGHGIHHCLGSFLARLEAKVFLELFFD 385

Qy        385 RFPEMKLKETPVFGY 399
              | : :::    |  |
Db        386 EFSDYRVEHDEVEFY 400


RESULT 15
T36526
probable cytochrome P450 hydroxylase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Feb-2001
C;Accession: T36526
R;Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999
A;Reference number: Z21609
A;Accession: T36526
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-411 <MUR>
A;Cross-references: EMBL:AL049754; PIDN:CAB42023.1; GSPDB:GN00070; SCOEDB:SCH10.14c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCH10.14c
C;Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: heme; iron; metalloprotein
F;250-385/Domain: cytochrome P450 homology <P45>
F;363/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match             20.2%;  Score 429;  DB 2;  Length 411;
  Best Local Similarity   27.1%;  Pred. No. 1.1e-22;
  Matches  110;  Conservative   93;  Mismatches  191;  Indels   12;  Gaps    7;

Qy         13 PAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERFAVSRE 72
              ||||  |: |  :  ||:||   ||      : |::    |::  :  ||||: || |     : :
Db         13 PAFD--PWDPAFVADPYPAFAELRARGRVLYYEPSDQWLVPHHADVSALLRDRRLGRTYQ 70

Qy         73 EWESSAEYSSAIP-----ELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQ 127
                :   ::      |           :  :|:  | | || |:|:||: :|| | :: |:  :
Db         71 HRFTHEDFGRTPPPPEQEPFHTLNDHGMLDLEPPDHTRIRRLVSKAFTPRTVERLKPYVH 130

Qy        128 RTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGL 187
                 | |:         |:: | || :|:  |: :| :| | |      |  ||    :   |
Db        131 GLADDLVARLVAAGGGDLLTDVAEPLPVAVIAEMLGIP-ESDRAPLRPWSAEICGM-YEL 188

Qy        188 VPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVAL 247
                |  :   | : ||: :     |   ::     ||: |    :|:::  |: |   :   ||: :|:::
Db        189 NPSEETAAKAVRASL-DFSDYLRALIAARRKEP-GDDLISGLIAAHDEDDRLTEQEMIST 246

Qy        248 VGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVR 307
                :: || : ||                 | | | |: |  ::|:  |: :|::|:|:|  |::   |
Db        247 CVLLLNAGHEATVNATTNGWLALFRHPDQLAALRADHSLVPSAVEELMRYDTPLQL-FER 305

Qy        308 FARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPG 367
                :  ::|  | :: :|  | :|  ||  |  ||: |:  |: |   :::   | | | |
Db        306 WVLDEIEIDGTTLPRGAEVAMLFGSANHDPAVFTDPERLDLTRRDNPHISFSAGIHYCIG 365

Qy        368 VSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVIL 413
               |||:|   :: :: :| | ::|    |    :    | :  | | |
Db        366 APLARIELAASMTSLLKRAPGLRLAAEPERRPNFVMRGLTELRVEL 411

Search completed: October  2, 2003, 17:35:09
Job time : 15.5739 secs
```

```
OM protein - protein search, using sw model

Run on:          October  2, 2003, 17:34:03 ; Search time 23.9103 Seconds
                                              (without alignments)
                                              2772.494 Million cell updates/sec

Title:           US-10-014-717-8
Perfect score:   2124
Sequence:        1 MTQEQANQSETKPAFDFKPF..........HPAFRNIESLNVILKPSKAG 419

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters:      587654

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
                 1:  /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                 2:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                 3:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                 4:  /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                 5:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                 6:  /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                 7:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                 8:  /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                 9:  /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
                 10:  /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
                 11:  /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
                 12:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                 13:  /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
                 14:  /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
                 15:  /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
                 16:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                 17:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                 18:  /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2124	100.0	419	14	US-10-014-717-8	Sequence 8, Appli
2	505.5	23.8	404	12	US-10-214-446-50	Sequence 50, Appl
3	498.5	23.5	392	12	US-10-214-446-32	Sequence 32, Appl
4	497.5	23.4	388	15	US-10-156-761-13776	Sequence 13776, A
5	495.5	23.3	408	12	US-10-214-446-4	Sequence 4, Appli
6	491	23.1	475	15	US-10-145-415-16	Sequence 16, Appl
7	488.5	23.0	475	15	US-10-145-415-22	Sequence 22, Appl
8	488	23.0	416	10	US-09-861-289-39	Sequence 39, Appl
9	488	23.0	416	10	US-09-860-846-39	Sequence 39, Appl
10	488	23.0	416	11	US-09-988-384B-39	Sequence 39, Appl
11	488	23.0	416	11	US-09-836-821-39	Sequence 39, Appl
12	486.5	22.9	470	15	US-10-145-415-6	Sequence 6, Appli
13	485.5	22.9	430	15	US-10-145-415-30	Sequence 30, Appl
14	484	22.8	416	11	US-09-793-708-18	Sequence 18, Appl
15	484	22.8	416	12	US-10-201-365-13	Sequence 13, Appl
16	484	22.8	416	12	US-10-160-539-18	Sequence 18, Appl
17	483.5	22.8	425	12	US-10-214-446-20	Sequence 20, Appl
18	483	22.7	399	15	US-10-156-761-9914	Sequence 9914, Ap
19	480.5	22.6	430	15	US-10-145-415-95	Sequence 95, Appl
20	475.5	22.4	430	15	US-10-145-415-8	Sequence 8, Appli
21	475.5	22.4	430	15	US-10-145-415-12	Sequence 12, Appl
22	475.5	22.4	430	15	US-10-145-415-20	Sequence 20, Appl
23	466	21.9	432	15	US-10-145-415-10	Sequence 10, Appl
24	463.5	21.8	430	15	US-10-145-415-32	Sequence 32, Appl
25	452.5	21.3	416	15	US-10-156-761-14997	Sequence 14997, A
26	452.5	21.3	430	15	US-10-145-415-18	Sequence 18, Appl
27	-451.5	21.3	429	15	US-10-145-415-14	Sequence 14, Appl
28	450	21.2	393	15	US-10-156-761-9525	Sequence 9525, Ap
29	450	21.2	405	15	US-10-156-761-12073	Sequence 12073, A
30	448	21.1	430	15	US-10-145-415-4	Sequence 4, Appli
31	448	21.1	430	15	US-10-145-415-28	Sequence 28, Appl
32	447	21.0	430	15	US-10-145-415-24	Sequence 24, Appl
33	447	21.0	457	15	US-10-156-761-11073	Sequence 11073, A
34	443.5	20.9	430	15	US-10-145-415-26	Sequence 26, Appl
35	443	20.9	399	12	US-10-314-657-37	Sequence 37, Appl
36	435.5	20.5	426	15	US-10-145-415-34	Sequence 34, Appl
37	435.5	20.5	430	15	US-10-145-415-2	Sequence 2, Appli
38	435	20.5	415	12	US-10-214-446-56	Sequence 56, Appl
39	433.5	20.4	399	15	US-10-156-761-7959	Sequence 7959, Ap
40	431	20.3	397	15	US-10-156-761-14715	Sequence 14715, A
41	428	20.2	415	12	US-10-214-446-46	Sequence 46, Appl
42	427.5	20.1	416	15	US-10-156-761-7958	Sequence 7958, Ap
43	425.5	20.0	402	15	US-10-205-032-8	Sequence 8, Appli
44	425	20.0	410	15	US-10-205-032-6	Sequence 6, Appli
45	418	19.7	404	12	US-10-214-446-40	Sequence 40, Appl

ALIGNMENTS

```
RESULT 1
US-10-014-717-8
; Sequence 8, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/10/014,717
;  CURRENT FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US/09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 419
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-10-014-717-8

  Query Match             100.0%;  Score 2124;  DB 14;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-198;
  Matches  419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60

Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
```

```
Db     121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy     181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240
Db     181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy     241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
Db     241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300

Qy     301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
Db     301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360

Qy     361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
Db     361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 2
US-10-214-446-50
; Sequence 50, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
;  APPLICANT: Weiner, David
;  APPLICANT:  Burk, Mark J.
;  APPLICANT:  Hitchman, Tim
;  APPLICANT:  Pujol, Catherine
;  APPLICANT:  Richardson, Toby
;  APPLICANT:  Short, Jay M.
;  TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
;  TITLE OF INVENTION:  THEM AND METHODS OF MAKING AND USING THEM
;  FILE REFERENCE: 09010-500001
;  CURRENT APPLICATION NUMBER: US/10/214,446
;  CURRENT FILING DATE:  2002-08-05
;  PRIOR APPLICATION NUMBER: US 60/309,497
;  PRIOR FILING DATE: 2001-08-03
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 50
;   LENGTH: 404
;   TYPE: PRT
;   ORGANISM: Bacterial
US-10-214-446-50

  Query Match             23.8%;  Score 505.5;  DB 12;  Length 404;
  Best Local Similarity   32.5%;  Pred. No. 1.4e-40;
  Matches  133;  Conservative   88;  Mismatches  163;  Indels   25;  Gaps   12;

Qy      22 PGYAEDPFPAIERLREA--TPIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAE 79
Db       5 PEFLRDPYPVYQRMRETGRMHLSSANTGRTWFLPHHADIRTALRDERFSASRKAGGFVNQ 64

Qy      80 YSSAI-PELS----DMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLL 134
Db      65 FPAEVRPEFARFNEAISRWIVLHDQPE-HRQLRQLMQQGFTRRLITTMEPKIQRVCDDLI 123

Qy     135 DA--RSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVD 192
Db     124 DAFVKRGSTEF--MTEYAHPFPAKVIAEMLGVNPEDYPAFVVWSEDLLNFAG-SLRP--- 177

Qy     193 EETKTLVASVTEGLALLHD----VLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALV 248
Db     178 --TLEMFRAAQDGLLAMMDYFARLLPERRENP-GDDLVSLLSAESEGEWMTAEQVLANC 234

Qy     249 GAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRF 308
Db     235 TQIIVAGHETTRNLVANGVELLLRYPEQRALLESRPELMPSAVREIMRFESPLQF-IRRV 293

Qy     309 ARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGV 368
Db     294 AREDFEFGGAEVREGDGLVLMLGSANRDPEAFDDPDTFDLTRNPTGHLAFGWGPHVCVGA 353

Qy     369 SLARLEAEIAVGTIFRRFPEMKLK-ETPVFGYHPAFRNIESLNVILKPS 416
Db     354 ALAELEGQVSFRTLLDRLPGLELRTHEPERIPNPMLRGFASLDLGFRES 402


RESULT 3
US-10-214-446-32
; Sequence 32, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
;  APPLICANT: Weiner, David
;  APPLICANT:  Burk, Mark J.
;  APPLICANT:  Hitchman, Tim
;  APPLICANT:  Pujol, Catherine
;  APPLICANT:  Richardson, Toby
;  APPLICANT:  Short, Jay M.
;  TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
;  TITLE OF INVENTION:  THEM AND METHODS OF MAKING AND USING THEM
;  FILE REFERENCE: 09010-500001
;  CURRENT APPLICATION NUMBER: US/10/214,446
;  CURRENT FILING DATE:  2002-08-05
;  PRIOR APPLICATION NUMBER: US 60/309,497
;  PRIOR FILING DATE: 2001-08-03
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
;   LENGTH: 392
;   TYPE: PRT
;   ORGANISM: Bacterial
US-10-214-446-32

  Query Match             23.5%;  Score 498.5;  DB 12;  Length 392;
  Best Local Similarity   33.8%;  Pred. No. 6.6e-40;
  Matches  127;  Conservative   64;  Mismatches  164;  Indels   21;  Gaps    5;

Qy      20 FAPGYAEDPFPAIERLREATPIF---YWDEGRSWVLTRYHDVSAVFRDERFAVSREEWES 76
Db       9 YGADFTANPYPYYAKLREAGPVHEVRMPDGFQFWLVVGHEEGRAALADPRLA-------K 61

Qy      77 SAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDA 136
Db      62 SPSVIGVRPPEEDIIGVHLLAADAPDHTRLRRLVTGEFTGRRVEGLRPRIQQLTTELADA 121

Qy     137 RSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETK 196
Db     122 MEPAGRADLVDAFAYPLPIIVICELLGVPAEDRDTFRRWSNQL-------VTPTGDQE-- 172

Qy     197 TLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEA-DGSRLSTKELVALVGAIIAAG 255
Db     173 -FGQAMVDFAAYLDALIEDKRAAGPTDDLLSALITARAEDGDRLSGPELRAMAYLLLIAG 231

Qy     256 TDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEY 315
Db     232 HETTVNLIANTVRNLLTHPEQLAALRADPDLLDGTIEESLRYDGPVETGTFRFTREAVTI 291

Qy     316 CGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLEA 375
Db     292 GGREIAAGQYVLVGIGALDRDPARFPDPDRFDIRRDTRGHLAFGHGIHYCLGAPLARLEG 351

Qy     376 EIAVGTIFRRFPEMKL 391
Db     352 RIALRTLLDRFPDLEL 367


RESULT 4
US-10-156-761-13776
; Sequence 13776, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
```

```
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13776
;   LENGTH: 388
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-13776

  Query Match             23.4%;  Score 497.5;  DB 15;  Length 388;
  Best Local Similarity   34.0%;  Pred. No. 8.1e-40;
  Matches  136;  Conservative   59;  Mismatches  144;  Indels   61;  Gaps   10;

Qy         21 APGYAEDPFPAIERLREATPIFY---WDEGRSWVLTRYHDVSAVFRDERFAVSREEWESS 77
              | |    ||:   :|||: |:        |     :|::||| ||              ||
Db          4 AEGLLADPYAVYDRLRDTAPVHRIAGTDGKPAWLVTRYDDV-----------RE----- 46

Qy         78 AEYSSAIPELSDMKKYGLFG------LPP----------EDHARVRKLVNPSFTSRAID 120
                  | | ||  ||: | |      |||          || |:|:|| :|| | ::
Db         47 ---GLANPLLSLDKKHALPGNYRGLALPPALDANLLNMDAPDHTRIRRLVGRAFTLRRVE 103

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
                || :: |  :||||       |::  ||  :|:   | ||| ||    || | |
Db        104 QLREPVRETAHRLLDALGTHGSTDLIASYAAPLPITVICDLLGVP---DEHRRDF----- 155

Qy        181 RALGVGLV------PQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEA 234
              ||    ||      | | |      ||   |        :| ::|:|| |:|: |:  :
Db        156 RAWTDPLVTPDPARPDVARE------SVVSLLGFFTGLLADKRKNPAD-DLLSDLIAVQE 208

Qy        235 DGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEV 294
              :| ||:  ||::|     |: || : |::|| ||| |||  || |      :: :|   :: :|: |
Db        209 EGDRLTEDELMSLAFLILFAGYENTVHLIGNAVLALLRHPEQLAALREDPARLPDAVGEF 268

Qy        295 LRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGA 354
               |::     :   || :|:    | ::  || | | :  :| || :  |  ||  |: ||
Db        269 ARYEGPALLAIRRFPVRDVTIGGVTVPAGETVLLSLSAANRDPSRFPDPDRLDLGRDAAG 328

Qy        355 SLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKET 394
               || | | | | |  |||| |:|:   :  |||:: | ||
Db        329 HLALGHGVHYCLGAPLARLETEVALAALLERFPDLALAET 368

RESULT 5
US-10-214-446-4
; Sequence 4, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
;  APPLICANT: Weiner, David
;  APPLICANT:  Burk, Mark J.
;  APPLICANT:  Hitchman, Tim
;  APPLICANT:  Pujol, Catherine
;  APPLICANT:  Richardson, Toby
;  APPLICANT:  Short, Jay M.
;  TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
;  TITLE OF INVENTION:  THEM AND METHODS OF MAKING AND USING THEM
;  FILE REFERENCE: 09010-500001
;  CURRENT APPLICATION NUMBER: US/10/214,446
;  CURRENT FILING DATE:  2002-08-05
;  PRIOR APPLICATION NUMBER: US 60/309,497
;  PRIOR FILING DATE: 2001-08-03
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 408
;   TYPE: PRT
;   ORGANISM: Bacterial
US-10-214-446-4

  Query Match             23.3%;  Score 495.5;  DB 12;  Length 408;
  Best Local Similarity   32.3%;  Pred. No. 1.4e-39;
  Matches  122;  Conservative   72;  Mismatches  157;  Indels   27;  Gaps    8;

Qy         27 DPFPAIERLRE--------ATPIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSA 78
              :|:|  ||:||          |           |::| | :   |  | | : |      :
Db         21 NPYPLFERIREHGVVQRVRLNPTL-----EVWMVTGYDEAVAALTDPRLS-SSPVGVNGL 74

Qy         79 EYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARS 138
                |   |  | :::      :       ||| |:| ||: :||:| ::  |  :|    |  ||| :
Db         75 EEEMAHQERTNVLMASMLVANGEDHTRLRNLVSKAFTARRVEQLAPRVQAHTDAFLDAVA 134

Qy        139 GQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLV--PQVDEETK 196
                :    |:| ::|  :||  :| |: :|||     ||          | |||:  |   |
Db        135 ARGSADLVSEFALPLPMAVLSDLIGIPAEGQPDPAR--------LAVGLIMPPNTPERLA 186

Qy        197 TLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGT 256
                 :  |        :: :|::  | ::|:|: |  |:|:  |:| :|| |:    :  ||
Db        187 KGARARAELTEFFEPLIAQRKKEP-KDDLLSALCAAQAE-ERISDRELTAMTILLTLAGH 244

Qy        257 DTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEYC 316
              :||  |||   |  |||  ||     :: :|  |:  |::|:||::   :  |  ||     |
Db        245 ETTASLIANGVHALLRHPEQFATLRDDPSLLPGAIEELLRYEGPVSRGVARFTTDPYEIG 304

Qy        317 GASIKKGEMVFLLIPSALRDGTVFSRPDVFDV-RRDTGASLAYGRGPHVCPGVSLARLEA 375
                | ::  |||: : : :| ||     :  |||: || ||:      ||:| | | | |  ||| ||
Db        305 GVTVPAGEMIIIGLAAANRDPARYDRPDILDVARREVPQQLAFGHGVHFCLGAPLARAEA 364

Qy        376 EIAVGTIFRRFPEMKLKE 393
                ||:||: ||||:::| :
Db        365 RIAIGTLLRRFPDLRLAD 382

RESULT 6
US-10-145-415-16
; Sequence 16, Application US/10145415
; Publication No. US20030068788A1
; GENERAL INFORMATION:
;  APPLICANT: Buckel, Thomas Gunter
;  APPLICANT:  Hammer, Philip Eugene
;  APPLICANT:  Hill, Dwight Steven
;  APPLICANT:  Ligon, James Madison
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Pachlatko, Johannes Paul
;  APPLICANT:  Zirkle, Ross Eric
;  TITLE OF INVENTION: Methods and Compositions for Making Emamectin
;  FILE REFERENCE: SYN-117 109846.312
;  CURRENT APPLICATION NUMBER: US/10/145,415
;  CURRENT FILING DATE:  2002-05-14
;  PRIOR APPLICATION NUMBER: US 60/291,149
;  PRIOR FILING DATE: 2001-05-16
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 475
;   TYPE: PRT
;   ORGANISM: Streptomyces albofaciens
US-10-145-415-16

  Query Match             23.1%;  Score 491;  DB 15;  Length 475;
  Best Local Similarity   29.6%;  Pred. No. 4.8e-39;
  Matches  129;  Conservative   83;  Mismatches  156;  Indels   68;  Gaps   11;

Qy          6 ANQSETKPAFDFKPFAPGYA---------------EDPFPAIERLREATPIF---YWDEG 47
               |:    | ||    :    | |                 ::|:      |||  |:     : |:
Db         34 ASPDTTSPATTDRTTLPSYVGLHPGEPNLMEPELLDNPYTGYGTLREQAPLVRARFIDDS 93

Qy         48 RSWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAIPELS-----DMKKYGLFGLP--- 99
```

```
                 |::||:  |  | ||:||                      : :| :          :   |||:|
Db         94 PIWLVTRFDVVREVMRDQRF----------VNNPTLVPGIGADQDPRARLIELFGIPEDL 143

Qy        100 ------------PEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEE---FD 144
                          | || |:|:||: :||:|  |  ||  ::|  |:||          |   |
Db        144 APYLTDTILTSDPPDHTRLRRLVSRAFTARRIQDLRPRVERITDELLARLPDHAEDGVVD 203

Qy        145 VVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLVASVTE 204
               :|  :|  :|:  |   |:  :     :||||:  |             |  | :: |
Db        204 LVEHFAYPLPITVICELVGIDEEDRALWRRFGADLA-----SLNP------KRIGATMPE 252

Qy        205 GLALLHDVLDERRRNPLENDVLTMLLQA-EADGSRLSTKELVALVGAIIAAGTDTTIYLI 263
               :| :|:|:|||  : |:|:|: |::| : |||  |:| ||  :: || :|| :||
Db        253 MIAHIHEVIDERRAD-LRDDLLSGLIRAQDDDGGRLSDVEMVTLVLTLVLAGHETTAHLI 311

Qy        264 AFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEYCGASIKKG 323
               :   | ||  |:    |:   :| |:   |: |::|:   ::    :|:| :| |   | :::|
Db        312 SNGTLALLTHPDQRRLIDEDPALLPRAVHELMRWCGPIQATQLRYAMEDTEVAGVQVRQG 371

Qy        324 EMVFLLIPSALRDGTVFSRPDVFDVRRDTGA----SLAYGRGPHVCPGVSLARLEAEIAV 379
               | :  : :|  |    :: |:  |: |          : :| | | | | |||| |||:|
Db        372 EALMFSLVAANHDPRHYTGPERLDLTRQPAGRAEDHVGFGHGMHYCLGASLARQEAEVAY 431

Qy        380 GTIFRRFPEMKLKETP 395
               | :  |:|:: |  ||
Db        432 GKLLTRYPDLALALTP 447
```

RESULT 7
US-10-145-415-22
; Sequence 22, Application US/10145415
; Publication No. US20030068788A1
; GENERAL INFORMATION:
; APPLICANT: Buckel, Thomas Gunter
; APPLICANT: Hammer, Philip Eugene
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ligon, James Madison
; APPLICANT: Molnar, Istvan
; APPLICANT: Pachlatko, Johannes Paul
; APPLICANT: Zirkle, Ross Eric
; TITLE OF INVENTION: Methods and Compositions for Making Emamectin
; FILE REFERENCE: SYN-117 109846.312
; CURRENT APPLICATION NUMBER: US/10/145,415
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 60/291,149
; PRIOR FILING DATE: 2001-05-16
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Streptomyces
US-10-145-415-22

```
  Query Match             23.0%;  Score 488.5;  DB 15;  Length 475;
  Best Local Similarity   29.7%;  Pred. No. 8.5e-39;
  Matches  131;  Conservative   83;  Mismatches  158;  Indels   69;  Gaps   12;

Qy          2 TQEQANQSETK-PAFDFKPFAPGYA---------------EDPFPAIERLREATPIF--- 42
               |   |  |:|  ||   :   | |                ::|:     |||  |:
Db         29 TPSAATSSDTTYPATTDRTTLPSYVGLHPGEPNLMEPELLDNPYTGYGTLREQAPLVRAR 88

Qy         43 YWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAIPELSDMKK-----YGLFG 97
               : |:    |::||:  |   | ||:||               :  :|  :    |        |||
Db         89 FIDDSPIWLVTRFDVVREVMRDQRF----------VNNPTLVPGIGADKDPRARLIELFG 138

Qy         98 LP---------------PEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEE 142
              :|               | || |:|:||: :||:|  |  ||  :::  | ||:      |
Db        139 IPEDLTPYLADTILTSDPPDHTRLRRLVSRAFTARRIQDLRPRVEQITDALLERLPDHAE 198

Qy        143 ---FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLV 199
                 |:|  :| :|:  |   |: :   |      :||||:  |         | |     |  :
Db        199 DGVVDLVEHFAYPLPITVICELVGIDEEDRTLWRRFGADLA-----SLNP------KRIG 247

Qy        200 ASVTEGLALLHDVLDERRRNPLENDVLTMLLQA-EADGSRLSTKELVALVGAIIAAGTDT 258
              |:: | :| :|:|:||||  | :|:|: |::| : ||  |||  |:| ||  :: || :|
Db        248 ATMPEMIAHIHEVIDERRA-ALRDDLLSGLIRAQDDDGGRLSDVEMVTLVLTLVLAGHET 306

Qy        259 TIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEYCGA 318
               | :||:   | ||  |:    |:   :| |:   |: |::|:   ::    :|:| :| |   |
Db        307 TAHLISNGTLALLTHPDQRRLIDEDPALLPRAVHELMRWCGPIQATQLRYAMEDTEVAGV 366

Qy        319 SIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGA----SLAYGRGPHVCPGVSLARLE 374
               :::|| :   : :|  |    :: |:  |: |          : :| | | | | |||| |
Db        367 QVRQGEALMFSLVAANHDPRHYTGPERLDLTRQPAGRAEDHVGFGHGMHYCLGASLARQE 426

Qy        375 AEIAVGTIFRRFPEMKLKETP 395
              ||:| | :  |:|::| ||
Db        427 AEVAYGKLLTRYPDLELALTP 447
```

RESULT 8
US-09-861-289-39
; Sequence 39, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-39

```
  Query Match             23.0%;  Score 488;  DB 10;  Length 416;
  Best Local Similarity   30.5%;  Pred. No. 7.7e-39;
  Matches  128;  Conservative   82;  Mismatches  185;  Indels   24;  Gaps    8;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHD 57
              | :|   : :  |  |       :| ||:|  |||   |        ||  |:: |
Db          1 MRRTQQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDR 60

Qy         58 VSAVFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSR 117
               ||  | ||:    ::| :|    :: : |         : :      |  | |:||||  || |
Db         61 ARAVLADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMR 114

Qy        118 AIDLLRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRF 175
              ::|||  :| || |:||     :    |::   |  :|:  || || ||       || :
Db        115 RVELLRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVW 174

Qy        176 GSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEAD 235
               |         : |    : :| :| :: |: |   :| :|   |: :  :: :: |
Db        175 TDAF-------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDED 224

Qy        236 GSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVL 295
              ||||:::||: :     :: || :||: |||    :  ||   |:  |   ::|:   |:  |::|:|
Db        225 GSRLTSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEML 284

Qy        296 RFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGAS 355
              |::  :     | ||  : ::   |   | |:  | :::  | |       |     ||:|||
Db        285 RYEGPVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGH 344

Qy        356 LAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
              ||:| | | | |  |||||| |||  :   | |::  |  :|      |:  :|  | :::|·:
Db        345 LAFGHGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403
```

```
RESULT 9
US-09-860-846-39
; Sequence 39, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 39
;   LENGTH: 416
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-39

  Query Match             23.0%;  Score 488;  DB 10;  Length 416;
  Best Local Similarity   30.5%;  Pred. No. 7.7e-39;
  Matches  128;  Conservative   82;  Mismatches  185;  Indels   24;  Gaps    8;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHD 57
              | : |     : : | |          :| ||:|   |||   |          ||  |:: |
Db          1 MRRTQQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDR 60

Qy         58 VSAVFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSR 117
                 ||  | ||:    ::| :|    :: : |          : :     |  | |:||||   || |
Db         61 ARAVLADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMR 114

Qy        118 AIDLLRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRF 175
                ::|||  :|  || |:||      :     |::    |   :|:  || || ||       || :
Db        115 RVELLRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVW 174

Qy        176 GSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEAD 235
                 |          : |     : :| :| ::   |: |    :| :|     |: :  ::  :: |
Db        175 TDAF-------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDED 224

Qy        236 GSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVL 295
              ||||:::||: :    :: || :||: |||   :   ||    |: |    ::|:   |:  |::|:|
Db        225 GSRLTSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEML 284

Qy        296 RFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGAS 355
              |::   :    | ||   : ::   |    |   |: | :::   | |       |   |   ||:||||
Db        285 RYEGPVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGH 344

Qy        356 LAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
              ||:|  |  |  |  |   ||||| |||    :    |  |::  |    :|    |:   :|   |  :::|  :
Db        345 LAFGHGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403

RESULT 10
US-09-988-384B-39
; Sequence 39, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 39
;   LENGTH: 416
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-39

  Query Match             23.0%;  Score 488;  DB 11;  Length 416;
  Best Local Similarity   30.5%;  Pred. No. 7.7e-39;
  Matches  128;  Conservative   82;  Mismatches  185;  Indels   24;  Gaps    8;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHD 57
              | : |     : : | |          :| ||:|   |||   |          ||  |:: |
Db          1 MRRTQQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDR 60

Qy         58 VSAVFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSR 117
                 ||  | ||:    ::| :|    :: : |          : :     |  | |:||||   || |
Db         61 ARAVLADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMR 114

Qy        118 AIDLLRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRF 175
                ::|||  :|  || |:||      :     |::    |   :|:  || || ||       || :
Db        115 RVELLRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVW 174

Qy        176 GSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEAD 235
                 |          : |     : :| :| ::   |: |    :| :|     |: :  ::  :: |
Db        175 TDAF-------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDED 224

Qy        236 GSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVL 295
              ||||:::||: :    :: || :||: |||   :   ||    |: |    ::|:   |:  |::|:|
Db        225 GSRLTSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEML 284

Qy        296 RFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGAS 355
              |::   :    | ||   : ::   |    |   |: | :::   | |       |   |   ||:||||
Db        285 RYEGPVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGH 344

Qy        356 LAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
              ||:|  |  |  |  |   ||||| |||    :    |  |::  |    :|    |:   :|   |  :::|  :
Db        345 LAFGHGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403

RESULT 11
US-09-836-821-39
; Sequence 39, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 39
;   LENGTH: 416
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-39

  Query Match             23.0%;  Score 488;  DB 11;  Length 416;
  Best Local Similarity   30.5%;  Pred. No. 7.7e-39;
  Matches  128;  Conservative   82;  Mismatches  185;  Indels   24;  Gaps    8;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHD 57
              | : |     : : | |          :| ||:|   |||   |          ||  |:: |
Db          1 MRRTQQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDR 60

Qy         58 VSAVFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSR 117
```

```
              || | ||:   ::| :|   :: : |        : :     | | |:|||| || |
Db        61 ARAVLADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMR 114

Qy       118 AIDLLRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRF 175
              ::||| :| || |:||    :   |::  |  :|: || || ||        || :
Db       115 RVELLRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVW 174

Qy       176 GSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEAD 235
              |       :  : :| :| :| : :| :  :|: :|  |: : :: : : |
Db       175 TDAF-------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDED 224

Qy       236 GSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVL 295
              ||||:::||: :.  :: || :||: |||  :  ||  |: |  ::|:  |:  |::|:|
Db       225 GSRLTSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEML 284

Qy       296 RFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGAS 355
              |::  :  | ||  : ::  |  |  |: | :::  ||     |  |  ||:||||
Db       285 RYEGPVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGH 344

Qy       356 LAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
              ||:| | | | |  ||||| |||  :  | |:: |  :|   |:  :|  | :::| :
Db       345 LAFGHGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403
```

RESULT 12
US-10-145-415-6
; Sequence 6, Application US/10145415
; Publication No. US20030068788A1
; GENERAL INFORMATION:
; APPLICANT: Buckel, Thomas Gunter
; APPLICANT: Hammer, Philip Eugene
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ligon, James Madison
; APPLICANT: Molnar, Istvan
; APPLICANT: Pachlatko, Johannes Paul
; APPLICANT: Zirkle, Ross Eric
; TITLE OF INVENTION: Methods and Compositions for Making Emamectin
; FILE REFERENCE: SYN-117 109846.312
; CURRENT APPLICATION NUMBER: US/10/145,415
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 60/291,149
; PRIOR FILING DATE: 2001-05-16
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Streptomyces rimosus
US-10-145-415-6

```
  Query Match             22.9%;  Score 486.5;  DB 15;  Length 470;
  Best Local Similarity   30.6%;  Pred. No. 1.3e-38;
  Matches  124;  Conservative   81;  Mismatches  147;  Indels   53;  Gaps   10;

Qy        22 PGYAEDPFPAIERLREATPIF---YWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSA 78
              |   |:|:      |||  |:     : |:   |::||:  |  | ||:||
Db        60 PELLENPYTGYGTLREQAPLVRARFIDDSPIWLVTRFDVVREVMRDQRF----------V 109

Qy        79 EYSSAIPELSDMKK-----YGLFGLP---------------PEDHARVRKLVNPSFTSRA 118
              : :| :    |          ||||:|               | || |:|:||: :||:|
Db       110 NNPTLVPGIGADKDPRARLIELFGIPEDLAPYLTDNILTSDPPDHTRLRRLVSRAFTARR 169

Qy       119 IDLLRAEIQRTVDQLLDARSGQEE---FDVVRDYAEGIPMRAISALLKVPAECDEKFRRF 175
              | ||  ::|  |:||:          |:|  :|  :|:  |  |  :  :         :|||
Db       170 IQDLRPRVERITDELLERLPDHAEDGVVDLVEHFAYPLPITVICELVGIDEEDRALWRRF 229

Qy       176 GSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQA-EA 234
              |:  |.      | |       | : |:: | :: :|:::||||  | :|:|: |::| :
Db       230 GADLA-----SLNP------KRIGATMPEMISHIHELIDERRA-ALRDDLLSGLIRAQDD 277

Qy       235 DGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEV 294
              || ||| |:| ||  :: || :|| :|.|:  ||| |:    |:  :| |:  |: |:
Db       278 DGGRLSDVEMVTLVLTLVLAGHETTAHLISNGTLALLTHPDQRRLIDEDPALLPRAVHEL 337

Qy       295 LRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGA 354
              :|:  ::   :|:|:|  | |  | :::|| :    : :| |      :: |: |: |
Db       338 MRWCGPIQATQLRYALEDTEVAGVQVRQGEALMFSLVAANHDPRHYTGPERLDLTRQPAG 397

Qy       355 ----SLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP 395
                  : :| | | | | |||| |||:| |  :  |:|:: |  ||
Db       398 RAEDHVGFGHGMHYCLGASLARQEAEVAYGKLLTRYPDLALALTP 442
```

RESULT 13
US-10-145-415-30
; Sequence 30, Application US/10145415
; Publication No. US20030068788A1
; GENERAL INFORMATION:
; APPLICANT: Buckel, Thomas Gunter
; APPLICANT: Hammer, Philip Eugene
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ligon, James Madison
; APPLICANT: Molnar, Istvan
; APPLICANT: Pachlatko, Johannes Paul
; APPLICANT: Zirkle, Ross Eric
; TITLE OF INVENTION: Methods and Compositions for Making Emamectin
; FILE REFERENCE: SYN-117 109846.312
; CURRENT APPLICATION NUMBER: US/10/145,415
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 60/291,149
; PRIOR FILING DATE: 2001-05-16
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Streptomyces lydicus
US-10-145-415-30

```
  Query Match             22.9%;  Score 485.5;  DB 15;  Length 430;
  Best Local Similarity   31.4%;  Pred. No. 1.4e-38;
  Matches  129;  Conservative   81;  Mismatches  150;  Indels   51;  Gaps   13;

Qy        22 PGYAEDPFPAIERLREATPIF---YWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSA 78
              |   |||      |||  |:      : |:   | :||: :|  | ||:||    |   |||
Db        25 PALIGDPFAGYGALREQGPVVRGRFVDDSPVWFVTRFEEVREVLRDQRF---RNNPVSSA 81

Qy        79 EYSSAIPELSDMKK-YGLFGLPPE---------------DHARVRKLVNPSFTSRAIDLL 122
              | || : : :    : | |                 || |:|:||: :||:| |  |
Db        82 --PDADPEDTPLSRLMDMMGFPEHLRVYLLGSILNNDAPDHTRLRRLVSRAFTARKITDL 139

Qy       123 RAEIQRTVDQLLDARSGQEE---FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179
              | : :  |:||    |    |:::  :|  :|:  |  |: :|  |    ::|  :|:
Db       140 RPRVAQIADELLARLPEHAEDGVVDLIQHFAYPLPITVICELVGIPEEDRPQWRTWGADL 199

Qy       180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQA-EADGSR 238
              | | |   | :::   | :       :|:::  |||   | :|:|: |::  : ||||
Db       200 -----VSLQP--DRMSRSFPAMIDH----IHELIAARRR-ALTDDLLSELIRTHDDDGSR 247

Qy       239 LSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFD 298
              || |:| :|  :: || :|| :||     ||  |:  | |:| :| |:  |:  |: |::|:
Db       248 LSDVEMVTMVLTVVLAGHETTAHLIGNGTAALLTHPDQLRLLKDDPALLPRAVHELMRWC 307

Qy       299 NILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGA---- 354
              : :  :|:| :|:| |  |:||: | |:: || ||    :: || |: |
Db       308 GPVHMTQLRYAAEDVELAGVRIRKGDAVQLILVSANRDPRHYTEPDRLDLTRHPAGHAEN 367

Qy       355 SLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMK-------LKETPVFG 398
              : :| | | | | :|| | |:|:|·: | |||:           |: ||| |
Db       368 HVGFGHGAHYCLGATLAKQEGEVALGALLRHFPELSLAVAPDALERTPVPG 418
```

RESULT 14
US-09-793-708-18
; Sequence 18, Application US/09793708
; Publication No. US20030104597A1

```
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 416
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-18

  Query Match             22.8%;  Score 484;  DB 11;  Length 416;
  Best Local Similarity   30.6%;  Pred. No. 1.9e-38;
  Matches  127;  Conservative   81;  Mismatches  183;  Indels   24;  Gaps    8;

Qy          5 QANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHDVSAV 61
Db          5 QQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDRARAV 64

Qy         62 FRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDL 121
Db         65 LADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMRRVEL 118

Qy        122 LRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179
Db        119 LRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVWTDAF 178

Qy        180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRL 239
Db        179 -------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDEDGSRL 228

Qy        240 STKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDN 299
Db        229 TSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEMLRYEG 288

Qy        300 ILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYG 359
Db        289 PVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGHLAFG 348

Qy        360 RGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
Db        349 HGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403
```

RESULT 15

```
US-10-201-365-13
; Sequence 13, Application US/10201365
; Publication No. US20030148469A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary
;  APPLICANT:  MCDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: COMBINATORIAL POLYKETIDE LIBRARIES PRODUCED USING A MODULAR
;  TITLE OF INVENTION:  PKS GENE CLUSTER AS SCAFFOLD
;  FILE REFERENCE: 300622002103
;  CURRENT APPLICATION NUMBER: US/10/201,365
;  CURRENT FILING DATE:  2002-07-22
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 416
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-10-201-365-13

  Query Match             22.8%;  Score 484;  DB 12;  Length 416;
  Best Local Similarity   30.6%;  Pred. No. 1.9e-38;
  Matches  127;  Conservative   81;  Mismatches  183;  Indels   24;  Gaps    8;

Qy          5 QANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHDVSAV 61
Db          5 QQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDRARAV 64

Qy         62 FRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDL 121
Db         65 LADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMRRVEL 118

Qy        122 LRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179
Db        119 LRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVWTDAF 178

Qy        180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRL 239
Db        179 -------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDEDGSRL 228

Qy        240 STKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDN 299
Db        229 TSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEMLRYEG 288

Qy        300 ILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYG 359
Db        289 PVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGHLAFG 348

Qy        360 RGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
Db        349 HGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403
```

Search completed: October 2, 2003, 17:47:43
Job time : 25.9103 secs

OM protein - protein search, using sw model

```
Run on:          October  2, 2003, 17:27:33 ; Search time 10.2473 Seconds
                                              (without alignments)
                                              1730.044 Million cell updates/sec

Title:           US-10-014-717-8
Perfect score:   2124
Sequence:        1 MTQEQANQSETKPAFDFKPF..........HPAFRNIESLNVILKPSKAG 419

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters:      328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                 1:  /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
                 2:  /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
                 3:  /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
                 4:  /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
                 5:  /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                 6:  /cgn2_6/ptodata/1/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2124	100.0	419	3	US-09-335-409-8	Sequence 8, Appli
2	2124	100.0	419	3	US-09-413-814-71	Sequence 71, Appl
3	2124	100.0	419	4	US-09-568-102-8	Sequence 8, Appli
4	2124	100.0	419	4	US-09-567-969-8	Sequence 8, Appli
5	2124	100.0	419	4	US-09-568-480-8	Sequence 8, Appli
6	2124	100.0	419	4	US-09-568-486-8	Sequence 8, Appli
7	2124	100.0	419	4	US-09-568-472-8	Sequence 8, Appli
8	2124	100.0	419	4	US-09-567-899-8	Sequence 8, Appli
9	488	23.0	416	3	US-09-105-537-39	Sequence 39, Appl
10	484.5	22.8	437	4	US-09-252-991A-17836	Sequence 17836, A
11	484	22.8	416	3	US-09-320-878-18	Sequence 18, Appl
12	484	22.8	416	4	US-09-141-908-13	Sequence 13, Appl
13	484	22.8	416	4	US-09-657-440-18	Sequence 18, Appl
14	459	21.6	403	6	5212296-9	Patent No. 5212296
15	421.5	19.8	468	4	US-09-252-991A-32437	Sequence 32437, A
16	371	17.5	409	3	US-09-385-028-12	Sequence 12, Appl
17	371	17.5	409	4	US-09-726-614-12	Sequence 12, Appl
18	370.5	17.4	399	4	US-08-765-907A-10	Sequence 10, Appl
19	361.5	17.0	412	1	US-08-102-863-11	Sequence 11, Appl
20	361.5	17.0	412	5	PCT-US92-10885-11	Sequence 11, Appl
21	354.5	16.7	406	6	5212296-6	Patent No. 5212296
22	325.5	15.3	395	4	US-09-266-965-129	Sequence 129, App
23	305	14.4	422	1	US-08-396-218-2	Sequence 2, Appli
24	305	14.4	422	1	US-08-760-116-2	Sequence 2, Appli
25	300	14.1	422	2	US-09-096-982-5	Sequence 5, Appli
26	300	14.1	422	2	US-08-653-650A-5	Sequence 5, Appli
27	300	14.1	474	2	US-09-096-982-8	Sequence 8, Appli
28	300	14.1	474	2	US-08-653-650A-8	Sequence 8, Appli
29	298	14.0	443	2	US-09-096-982-9	Sequence 9, Appli
30	298	14.0	443	2	US-08-653-650A-9	Sequence 9, Appli
31	262.5	12.4	189	4	US-09-679-279-20	Sequence 20, Appl
32	257	12.1	382	3	US-09-320-878-7	Sequence 7, Appli
33	257	12.1	382	4	US-09-141-908-7	Sequence 7, Appli
34	257	12.1	382	4	US-09-657-440-7	Sequence 7, Appli
35	257	12.1	402	3	US-09-105-537-22	Sequence 22, Appl
36	257	12.1	3782	3	US-09-105-537-4	Sequence 4, Appli
37	185.5	8.7	472	2	US-08-622-166A-2	Sequence 2, Appli
38	185.5	8.7	472	2	US-08-622-166A-4	Sequence 4, Appli
39	184	8.7	520	4	US-09-527-073-2	Sequence 2, Appli
40	175.5	8.3	377	4	US-09-679-279-3	Sequence 3, Appli
41	172.5	8.1	513	4	US-09-351-229-4	Sequence 4, Appli
42	170	8.0	444	3	US-09-413-814-95	Sequence 95, Appl
43	170	8.0	444	3	US-09-413-814-96	Sequence 96, Appl
44	164	7.7	504	4	US-09-126-420A-18	Sequence 18, Appl
45	163	7.7	495	1	US-08-532-065B-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-335-409-8
; Sequence 8, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 419
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-335-409-8

  Query Match             100.0%;  Score 2124;  DB 3;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-210;
  Matches  419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60

Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300

Qy        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
```

```
Qy        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 2
US-09-413-814-71
; Sequence 71, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
;  APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
;  APPLICANT: Bristol-Myers Squibb, Co.
;  APPLICANT: Beyer, Stefan
;  APPLICANT: Bloecker, Helmut
;  APPLICANT: Brandt, Petra
;  APPLICANT: Cino, Paul M
;  APPLICANT: Dougherty, Brian A
;  APPLICANT: Goldberg, Steven L
;  APPLICANT: Hofle, Gerhard
;  APPLICANT: Mueller, Joachim
;  APPLICANT: Reichenbach, Hans
;  TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
;  TITLE OF INVENTION: heteropolyketide compounds
;  FILE REFERENCE: PCT/US 99/23535
;  CURRENT APPLICATION NUMBER: US/09/413,814
;  CURRENT FILING DATE: 1999-10-07
;  EARLIER APPLICATION NUMBER: DE 198 46 493.2
;  EARLIER FILING DATE: 1998-10-09
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
;   LENGTH: 419
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-413-814-71

  Query Match             100.0%;  Score 2124;  DB 3;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-210;
  Matches  419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60

Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300

Qy        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360

Qy        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 3
US-09-568-102-8
; Sequence 8, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,102
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 419
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-568-102-8

  Query Match             100.0%;  Score 2124;  DB 4;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-210;
  Matches  419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60

Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300

Qy        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360

Qy        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 4
US-09-567-969-8
; Sequence 8, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 419
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-567-969-8

  Query Match             100.0%;  Score 2124;  DB 4;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-210;
  Matches  419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60

Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240.
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300

Qy        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360

Qy        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 5
US-09-568-480-8
; Sequence 8, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 419
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-568-480-8

  Query Match             100.0%;  Score 2124;  DB 4;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-210;
  Matches  419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60

Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300

Qy        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360

Qy        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 6
US-09-568-486-8
; Sequence 8, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 419
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-568-486-8

  Query Match             100.0%;  Score 2124;  DB 4;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-210;
  Matches  419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60

Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
```

```
Qy        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360

Qy        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 7
US-09-568-472-8
; Sequence 8, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 419
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-568-472-8

  Query Match             100.0%;  Score 2124;  DB 4;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-210;
  Matches  419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60

Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300

Qy        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360

Qy        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 8
US-09-567-899-8
; Sequence 8, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 419
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-567-899-8

  Query Match             100.0%;  Score 2124;  DB 4;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-210;
  Matches  419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60

Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300

Qy        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360

Qy        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 9
US-09-105-537-39
; Sequence 39, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 39
;   LENGTH: 416
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-39
```

```
  Query Match             23.0%;  Score 488;  DB 3;  Length 416;
  Best Local Similarity   30.5%;  Pred. No. 9.4e-42;
  Matches  128;  Conservative   82;  Mismatches  185;  Indels   24;  Gaps    8;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHD 57
Db          1 MRRTQQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDR 60

Qy         58 VSAVFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSR 117
Db         61 ARAVLADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMR 114

Qy        118 AIDLLRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRF 175
Db        115 RVELLRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVW 174

Qy        176 GSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEAD 235
Db        175 TDAF-------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDED 224

Qy        236 GSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVL 295
Db        225 GSRLTSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEML 284

Qy        296 RFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGAS 355
Db        285 RYEGPVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGH 344

Qy        356 LAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
Db        345 LAFGHGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403

RESULT 10
US-09-252-991A-17836
; Sequence 17836, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 17836
;   LENGTH: 437
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17836

  Query Match             22.8%;  Score 484.5;  DB 4;  Length 437;
  Best Local Similarity   33.5%;  Pred. No. 2.3e-41;
  Matches  116;  Conservative   74;  Mismatches  139;  Indels   17;  Gaps    5;

Qy         50 WVLTRYHDVSAVFRDERFAVSREEWESSAEYS----SAIPELSDMKKYGLFGLPPEDHAR 105
Db         60 WVVTRYRDARKVLNHP--GVRRDARQAAELYAKRTGSPRAGIGEGLSHHMLNLDPPDHTR 117

Qy        106 VRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVP 165
Db        118 LRSLVGRAFTPRQVERLQPHIERITEELLDAMAGREQADLMADFAIPLTIAVIFELLGIP 177

Qy        166 AECDEKFRRFGSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDV 225
Db        178 EAEREHARQSWERQAELLS-------PEEAQALADAQVDYLRVL---LEAKRRQPAD-DV 226

Qy        226 LTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPG 285
Db        227 YSGLVQAADESGQLSEAELVSMAHLLMMSGFETTMNMIGNALVTLLVNPEQLALLRAQPE 286

Qy        286 LMRNALDEVLRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDV 345
Db        287 LLPNAMEELVRHDSPVRASMLRFTVEDVELDGVTIPAGEYILVSNLTANHDAERFDDPDR 346

Qy        346 FDVRRDTGASLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKL 391
Db        347 LDLTRNTDGHLGYGFGVHYCVGASLARLEGRIAIQRLLARFPDLQL 392

RESULT 11
US-09-320-878-18
; Sequence 18, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 416
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-18

  Query Match             22.8%;  Score 484;  DB 3;  Length 416;
  Best Local Similarity   30.6%;  Pred. No. 2.4e-41;
  Matches  127;  Conservative   81;  Mismatches  183;  Indels   24;  Gaps    8;

Qy          5 QANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHDVSAV 61
Db          5 QQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDRARAV 64

Qy         62 FRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDL 121
Db         65 LADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMRRVEL 118

Qy        122 LRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179
Db        119 LRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVWTDAF 178

Qy        180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRL 239
Db        179 -------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDEDGSRL 228

Qy        240 STKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDN 299
Db        229 TSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEMLRYEG 288

Qy        300 ILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYG 359
Db        289 PVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGHLAFG 348
```

```
Qy        360 RGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
              | | | |  |||||| |||  :  | |:: |  :|   |:  :|  | :::| :
Db        349 HGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403
```

RESULT 12
US-09-141-908-13
; Sequence 13, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-141-908-13

```
  Query Match             22.8%;  Score 484;  DB 4;  Length 416;
  Best Local Similarity   30.6%;  Pred. No. 2.4e-41;
  Matches  127;  Conservative   81;  Mismatches  183;  Indels   24;  Gaps    8;

Qy          5 QANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHDVSAV 61
              |   : :  |  |           :| ||:|   |||    |        ||    |:: |   ||
Db          5 QQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDRARAV 64

Qy         62 FRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDL 121
                | ||:   ::| :|   :: : |         : :     |   |  | |:||||   || |  ::|
Db         65 LADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMRRVEL 118

Qy        122 LRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179
              ||   :|   || |:||          :     |::     | :|:  || || ||       || :   |
Db        119 LRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVWTDAF 178

Qy        180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRL 239
                       : |      :  :| :| ::  |: |     :| :|     |: :  ::  :: |||||
Db        179 -------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDEDGSRL 228

Qy        240 STKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDN 299
              :::||: :   :: || :||: |||   :  ||  |: |   ::|:  |:  |::|:||::
Db        229 TSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEMLRYEG 288

Qy        300 ILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYG 359
               :    |.||  : ::  |   |   |:  | :::   | |      |   |  ||:||||    ||:|
Db        289 PVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGHLAFG 348

Qy        360 RGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
              | | | |  |||||| |||  :  | |:: |  :|   |:  :|  | :::| :
Db        349 HGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403
```

RESULT 13
US-09-657-440-18
; Sequence 18, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-657-440-18

```
  Query Match             22.8%;  Score 484;  DB 4;  Length 416;
  Best Local Similarity   30.6%;  Pred. No. 2.4e-41;
  Matches  127;  Conservative   81;  Mismatches  183;  Indels   24;  Gaps    8;

Qy          5 QANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHDVSAV 61
              |   : :  |  |           :| ||:|   |||    |        ||    |:: |   ||
Db          5 QQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDRARAV 64

Qy         62 FRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDL 121
                | ||:   ::| :|   :: : |         : :     |   |  | |:||||   || |  ::|
Db         65 LADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMRRVEL 118

Qy        122 LRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179
              ||   :|   || |:||          :     |::     | :|:  || || ||       || :   |
Db        119 LRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVWTDAF 178

Qy        180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRL 239
                       : |      :  :| :| ::  |: |     :| :|     |: :  ::  :: |||||
Db        179 -------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDEDGSRL 228

Qy        240 STKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDN 299
              :::||: :   :: || :||: |||   :  ||  |: |   ::|:  |:  |::|:||::
Db        229 TSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEMLRYEG 288

Qy        300 ILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYG 359
               :    | ||  : ::  |   |   |:  | :::   | |      |   |  ||:||||    ||:|
Db        289 PVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGHLAFG 348

Qy        360 RGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
              | | | |  |||||| |||.  :  | |:: |  :|   |:  :|  | :::| :
Db        349 HGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403
```

RESULT 14
5212296-9
;Patent No. 5212296
; APPLICANT: DEAN, CAROLINE;HARDER, PATRICIA A.;LETO, KENNETH
;J.; O'KEEFE, DANIEL P.;OMER, CHARLES A.;ROMESSER, JAMES A.
;TEPPERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
;CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989

```
;SEQ ID NO:9:
;       LENGTH: 403
5212296-9

  Query Match             21.6%;  Score 459;  DB 6;  Length 403;
  Best Local Similarity   31.7%;  Pred. No. 8.8e-39;
  Matches  127;  Conservative   73;  Mismatches  157;  Indels   44;  Gaps   12;

Qy         21 APGYAEDPFPAIERLREATPI---FYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESS 77
Db         20 APGCPFDPAPDVTEAARTEPVTRATLWDGSSCWLVTRHQDVRAVLGDPRF--------SA 71

Qy         78 AEYSSAIPELSDMKKYGLFGLPP-------EDHARVRKLVNPSFTSRAIDLLRAEIQRTV 130
Db         72 DAHRTGFPFLTAGGR-EIIGTNPTFLRMDDPEHARLRRMLTADFIVKKVEAMRPEVQRLA 130

Qy        131 DQLLD-ARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVP 189
Db        131 DDLVDRMTTGRTSADLVTEFALPLPSLVICLLLGVPYE-DHAF--FQERSRVLLTLRSTP 187

Qy        190 QVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVG 249
Db        188 ---EEVRAAQDELLEYLARL---ARTKRERPDDAIISRLVARGELDDTQIATMGRLLLV- 240

Qy        250 AIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFA 309
Db        241 ----AGHETTANMTALSTLVLLRNPDQLARLRAEPALVKGAVEELLRYLTIVHNGVPRIA 296

Qy        310 RQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVS 369
Db        297 TEDVLIGGRTIAAGEGVLCMISSANRDAEVFPGGDDLDVARDARRHVAFGFGVHQCLGQP 356

Qy        370 LARLEAEIAVGTIFRRFPEMKL----KETP------VFGYH 400
Db        357 LARVELQIAIETLLRRLPDLRLAVPHEEIPFRGDMAIYGVH 397


RESULT 15
US-09-252-991A-32437
; Sequence 32437, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 32437
;   LENGTH: 468
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32437

  Query Match             19.8%;  Score 421.5;  DB 4;  Length 468;
  Best Local Similarity   29.5%;  Pred. No. 8.3e-35;
  Matches  133;  Conservative   70;  Mismatches  177;  Indels   71;  Gaps   15;

Qy          5 QANQSETKPA-------FDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHD 57
Db         36 QPRHDAQRPALAPRSDGFDIHTYHPDFVADPYPLLRLIRSRAPVCR-DQASIWWISRYAD 94

Qy         58 VSAVFRDERFAVS---------REEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRK 108
Db         95 VSACLRDRRFSADPARLGAAGVRQGGASWFGHQQLQPLARFYDNFMLFNDAPR-HTRLRR 153

Qy        109 LVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAEC 168
Db        154 LFAPAFGPDAVRRWEARIEVLVEELLDSLLERREPDLLRDFAEPLTIRVAAELFGFPRED 213

Qy        169 DEKFRRFGSATARALGVGLVPQVDEETKTLVAS------VTEGLALLHDVLDERRRN--- 219
Db        214 TGQLLPWG----RDLAAGL---------DLAASHGDAGQINRSAAAFSDYLQRQARGWSD 260

Qy        220 ----------PLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLN 269
Db        261 GSSRPPSGAAPSIFDGAAML---EAG---LGLEDLVAAYAMVFMAAFETTISMVGNATLA 314

Qy        270 LLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLL 329
Db        315 LLTHPDQLDLLRRCPELAANAVEELLRFDGAVR-GGVRCTLEEVEIGGQRIPPGEKVWLS 373

Qy        330 IPSALRDGTVFSRPDVFDVRR-DTGASLAYGRGPHVCPGVSLARLEAEIAV-GTIFRRF- 386
Db        374 FLAANRDPEMFAAPDRLQLQRANAKQHVAFAHGPHYCLGAYLARLELQCALRGLVRRRFA 433

Qy        387 -----PEMKLKETPVFGYHPAFRNIESLNVI 412
Db        434 LASEPTDLRWRRSSV------FRTLERLPIV 458

Search completed: October  2, 2003, 17:35:54
Job time : 12.2473 secs
```

```
OM protein - protein search, using sw model

Run on:          October  2, 2003, 17:27:29 ; Search time 28.237 Seconds
                                              (without alignments)
                                              2355.298 Million cell updates/sec

Title:           US-10-014-717-8
Perfect score:   2124
Sequence:        1 MTQEQANQSETKPAFDFKPF..........HPAFRNIESLNVILKPSKAG 419

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters:      1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_19Jun03:*
                 1:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
                 2:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
                 3:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
                 4:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
                 5:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
                 6:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
                 7:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
                 8:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
                 9:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
                 10:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
                 11:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
                 12:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
                 13:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
                 14:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
                 15:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
                 16:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
                 17:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
                 18:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
                 19:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
                 20:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
                 21:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
                 22:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
                 23:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
                 24:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2124	100.0	419	21	AAY58579	Sorangium cellulos
2	526.5	24.8	396	21	AAB15502	Bacillus subtilis
3	525.5	24.7	396	21	AAB15503	Bacillus subtilis
4	518.5	24.4	396	21	AAB15501	Bacillus subtilis
5	516	24.3	399	24	ABP57702	Saccharopolyspora
6	491	23.1	475	24	AAE34729	Streptomyces albof
7	488.5	23.0	475	24	AAE34732	Streptomyces rimos
8	486.5	22.9	470	24	AAE34724	Streptomyces rimos
9	485.5	22.9	430	24	AAE34736	Streptomyces lydic
10	484	22.8	416	21	AAB18654	Amino acid sequenc
11	484	22.8	416	21	AAY67218	Macrolide ring mod
12	484	22.8	416	24	ABG71678	S. venezuelae picr
13	480.5	22.6	430	24	AAE34780	Streptomyces tuber
14	476.5	22.4	415	21	AAY77196	S. venezuelae macr
15	475.5	22.4	430	24	AAE34725	Streptomyces lydic
16	475.5	22.4	430	24	AAE34727	Streptomyces chatt
17	475.5	22.4	430	24	AAE34731	Streptomyces kasug
18	466	21.9	432	24	AAE34726	Streptomyces speci
19	463.5	21.8	430	24	AAE34737	Streptomyces lydic
20	452.5	21.3	430	24	AAE34730	Streptomyces plate
21	452	21.3	482	23	ABB81316	Streptomyces sp. c
22	451.5	21.3	429	24	AAE34728	Streptomyces speci
23	450	21.2	396	23	ABB81315	Streptomyces sp. c
24	449	21.1	448	24	ABU11985	M. echinospora cal
25	448	21.1	430	24	AAE34723	Streptomyces tuber
26	448	21.1	430	24	AAE34735	Streptomyces plate
27	447	21.0	430	24	AAE34733	Streptomyces tuber
28	443.5	20.9	430	24	AAE34734	Streptomyces plate
29	443	20.9	399	24	ABU11376	Protein encoded by
30	435.5	20.5	426	24	AAE34738	Streptomyces tuber
31	435.5	20.5	430	24	AAE34722	Streptomyces tuber
32	432	20.3	403	12	AAR11350	Cytochrome enzyme
33	420	19.8	411	19	AAW54389	Actinomadura hibis
34	414.5	19.5	404	12	AAR14724	6-hydroxylase enco
35	413	19.4	396	15	AAR60777	Mycinamicin IV hyd
36	407.5	19.2	587	18	AAW33274	S. fradiae tylosin
37	406.5	19.1	407	24	ABP95990	Microtetraspora re
38	405.5	19.1	418	21	AAB13589	Streptomyces globi
39	403.5	19.0	420	21	AAY83790	S. fradiae tylosin
40	401.5	18.9	402	24	ABU11981	M. echinospora cal
41	386	18.2	398	22	AAB31560	Pimaricin biosynth
42	385.5	18.1	405	24	ABP95987	Streptomyces sp. T
43	382	18.0	406	24	ABP95992	Streptomyces livid
44	371	17.5	409	22	AAE07915	S. clavuligerus pr
45	370.5	17.4	398	17	AAW11585	Streptomyces prist

ALIGNMENTS

```
RESULT 1
AAY58579
ID   AAY58579 standard; Protein; 419 AA.
XX
AC   AAY58579;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone macrolactone oxidase EPOS F.
XX
KW   EPOS F; epothilone macrolactone oxidase; epothilone biosynthesis;
KW   taxol substitute; anticancer.
XX
OS   Sorangium cellulosum.
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   N-PSDB; AAZ55887.
```

```
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy  -
XX
PS   Claim 12; Page 158-159; 174pp; English.
XX
CC   This sequence represents a Sorangium cellulosum epothilone macrolactone
CC   oxidase, EPOS F, which is one of several epothilone biosynthetic
CC   enzymes encoded by a 68.75 kb contig. Epothilones A and B are
CC   16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence   419 AA;

  Query Match            100.0%;  Score 2124;  DB 21;  Length 419;
  Best Local Similarity  100.0%;  Pred. No. 6.8e-198;
  Matches  419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSA 60

Qy         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID 120

Qy        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATA 180

Qy        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS 240

Qy        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI 300

Qy        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGR 360

Qy        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419


RESULT 2
AAB15502
ID   AAB15502 standard; Protein; 396 AA.
XX
AC   AAB15502;
XX
DT   14-FEB-2001  (first entry)
XX
DE   Bacillus subtilis hydroxylating protein #2.
XX
KW   Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;
KW   hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
XX
OS   Bacillus subtilis.
XX
PN   WO200044886-A1.
XX
PD   03-AUG-2000.
XX
PF   28-JAN-2000; 2000WO-JP00472.
XX
PR   29-JAN-1999;   99JP-0021707.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Endo H,  Yonetani Y,  Mizoguchi H,  Hashimoto S,  Ozaki A;
XX
DR   WPI; 2000-548827/50.
DR   N-PSDB; AAA95667.
XX
PT   New protein derived from Bacillus genus microorganism useful for
PT   producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT   inhibitors  -
XX
PS   Claim 10; Page 94-96; 111pp; Japanese.
XX
CC   This sequence represents a novel Bacillus derived protein having
CC   the activity of producing a hydroxylated bicyclic compound or the
CC   corresponding lactone from a bicyclic compound or the corresponding
CC   lactone. The protein is used for preparing hydroxylated compounds
CC   useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase
CC   inhibitors and hypocholesterolaemic agents.
XX
SQ   Sequence   396 AA;

  Query Match            24.8%;  Score 526.5;  DB 21;  Length 396;
  Best Local Similarity  33.2%;  Pred. No. 2.6e-42;
  Matches  126;  Conservative   72;  Mismatches  153;  Indels   29;  Gaps    6;

Qy         25 AEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAI 84
              | ||| |:|:  |: : :| : |  | ||  |  |:                    :|| :
Db         23 AYHPFPWYESMRKDAPVSFDEENQVWSVFLYDDVKKVVGDKEL-----------FSSCM 70

Qy         85 PELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEEFD 144
              |: :      : : |  | ::| :|| :|| ||:          ||    |:|:    |: |||
Db         71 PQQTSSIGNSIINMDPPKHTKIRSVVNKAFTPRAMKQWEPRIQEITDELIQKFQGRSEFD 130

Qy        145 VVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLVA---- 200
              :| |:: :|:  || || ||:   |:|: :            | |          || :
Db        131 LVHDFSYPLPVIVISELLGVPSAHMEQFKAWSD-----LLVSTPKDKSEEAEKAFLEERD 185

Qy        201 SVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTI 260
              | ||          : | :|| | |::::|::|| |:|| :||:| |   :: || :||
Db        186 KCEEELAAFFAGIIEEKRNKPEQDIISILVEAEETGEKLSGEELIPLCTLLLVAGNETTT 245

Qy        261 YLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRF---DNILRIGTVRFARQDLEYCG 317
              ||: |: ::| :|   | ::: | ||  |::| |||     :||   | |::| |  |
Db        246 NLISNAMYSILETPGVYEELRSHPELMPQAVEEALRFRAPAPVLR----RIAKRDTEIGG 301

Qy        318 ASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLEAEI 377
              ||:|:||   : || ||    | || :|:||       :|:| | | | | ||||| |
Db        302 HLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHPNPHIAFGHGIHFCLGAPLARLEANI 361

Qy        378 AVGTIFRRFPEMK-LKETPV 396
              |: ::   || |: :  ||:
Db        362 ALTSLISAFPHMECVSITPI 381

RESULT 3
```

```
AAB15503
ID   AAB15503 standard; Protein; 396 AA.
XX
AC   AAB15503;
XX
DT   14-FEB-2001  (first entry)
XX
DE   Bacillus subtilis hydroxylating protein #3.
XX
KW   Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;
KW   hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
XX
OS   Bacillus subtilis.
XX
PN   WO200044886-A1.
XX
PD   03-AUG-2000.
XX
PF   28-JAN-2000; 2000WO-JP00472.
XX
PR   29-JAN-1999;    99JP-0021707.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Endo H,  Yonetani Y,  Mizoguchi H,  Hashimoto S,  Ozaki A;
XX
DR   WPI; 2000-548827/50.
DR   N-PSDB; AAA95669.
XX
PT   New protein derived from Bacillus genus microorganism useful for
PT   producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT   inhibitors  -
XX
PS   Claim 10; Page 105-107; 111pp; Japanese.
XX
CC   This sequence represents a novel Bacillus derived protein having
CC   the activity of producing a hydroxylated bicyclic compound or the
CC   corresponding lactone from a bicyclic compound or the corresponding
CC   lactone. The protein is used for preparing hydroxylated compounds
CC   useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase
CC   inhibitors and hypocholesterolaemic agents.
XX
SQ   Sequence   396 AA;

  Query Match             24.7%;  Score 525.5;  DB 21;  Length 396;
  Best Local Similarity   33.2%;  Pred. No. 3.2e-42;
  Matches  126;  Conservative   71;  Mismatches  154;  Indels   29;  Gaps    6;

Qy         25 AEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAI 84
               |  |||  |  :|:  |:  : :|  :  |  :  | ||  |  |:            :|| :
Db         23 AYHPFPWYESMRKDAPVSFDEENQVWSVFLYDDVKKVVGDKEL------------FSSCM 70

Qy         85 PELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEEFD 144
              |: :        :  :  |   | ::|  :||  :|| ||:         ||    |:|:     |: |||
Db         71 PQQTSSIGNSIISMDPPKHTKIRSVVNKAFTPRAMKQWEPRIQEITDELIQKFQGRSEFD 130

Qy        145 VVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLVA---- 200
              :| ||:  :|:  || || ||:    |:|: :           | |          || :
Db        131 LVHDYSYPLPVIVISELLGVPSAHMEQFKAWSD-----LLVSTPKDKSEEAEKAFLEERD 185

Qy        201 SVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTI 260
                     | ||       :  |  :||  |  |::::|::||   |  :||  :||:  |     ::  || :||
Db        186 KCEEELAAFFAGIIEEKRNKPEQDIISILVEAEETGEKLSGEELIPLCTLLLVAGNETTT 245

Qy        261 YLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRF---DNILRIGTVRFARQDLEYCG 317
                ||:  |:  ::|  :|      |  :::  |  ||     |::|  |||          :||       |  |::|  |   |
Db        246 NLISNAMFSILETPGVYEELRSHPELMPQAVEEALRFRAPAPVLR----RIAKRDTEIGG 301

Qy        318 ASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLEAEI 377
                |||:|:  |       :  ||  ||        |  |:|  :||:||          :|:|  |  |  |  |     ||||||   |
Db        302 HLIKEGDTVLAFVASANRDEAKFDRPHMFDIRRHPNPHIAFGHGIHFCLGAPLARLEANI 361

Qy        378 AVGTIFRRFPEMK-LKETPV 396
```

```
              |: ::    || |: :  ||:
Db        362 ALTSLISAFPHMECVSITPI 381

RESULT 4
AAB15501
ID   AAB15501 standard; Protein; 396 AA.
XX
AC   AAB15501;
XX
DT   14-FEB-2001  (first entry)
XX
DE   Bacillus subtilis hydroxylating protein.
XX
KW   Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;
KW   hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
XX
OS   Bacillus subtilis.
XX
PN   WO200044886-A1.
XX
PD   03-AUG-2000.
XX
PF   28-JAN-2000; 2000WO-JP00472.
XX
PR   29-JAN-1999;    99JP-0021707.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Endo H,  Yonetani Y,  Mizoguchi H,  Hashimoto S,  Ozaki A;
XX
DR   WPI; 2000-548827/50.
DR   N-PSDB; AAA95666.
XX
PT   New protein derived from Bacillus genus microorganism useful for
PT   producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT   inhibitors  -
XX
PS   Claim 8; Page 68-71; 111pp; Japanese.
XX
CC   This sequence represents a novel Bacillus derived protein having
CC   the activity of producing a hydroxylated bicyclic compound or the
CC   corresponding lactone from a bicyclic compound or the corresponding
CC   lactone. The protein is used for preparing hydroxylated compounds
CC   useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase
CC   inhibitors and hypocholesterolaemic agents.
XX
SQ   Sequence   396 AA;

  Query Match             24.4%;  Score 518.5;  DB 21;  Length 396;
  Best Local Similarity   32.6%;  Pred. No. 1.5e-41;
  Matches  124;  Conservative   72;  Mismatches  155;  Indels   29;  Gaps    6;

Qy         25 AEDPFPAIERLREATPIFYWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAI 84
               |  |||  |  :|:  |:  : :|  :  |  :  | ||  |  |:            :|| :
Db         23 AYHPFPWYESMRKDAPVSFDEENQVWSVFLYDDVKKVVGDKEL------------FSSCM 70

Qy         85 PELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEEFD 144
              |: :        :  :  |   | ::|  :||  :|| ||:         ||    |:|:     |: |||
Db         71 PQQTSSIGNSIINMDPPKHTKIRSVVNKAFTPRVMKQWEPRIQEITDELIQKFQGRSEFD 130

Qy        145 VVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLVA---- 200
              :| |::  :|:  || || ||:    |:|: :           | |          || :
Db        131 LVHDFSYPLPVIVISELLGVPSAHMEQFKAWSD-----LLVSTPKDKSEEAEKAFLEERD 185

Qy        201 SVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGAIIAAGTDTTI 260
                     | ||       :  |  :||  |  |::::|::||   |  :||  :||:  |     ::  || :||
Db        186 KCEEELAAFFAGIIEEKRNKPEQDIISILVEAEETGEKLSGEELIPFCTLLLVAGNETTT 245

Qy        261 YLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRF---DNILRIGTVRFARQDLEYCG 317
                ||:  |:  ::|  :|      |  :::  |  ||     |::|  |||          :||       |  |::|  |   |
Db        246 NLISNAMYSILETPGVYEELRSHPELMPQAVEEALRFRAPAPVLR----RIAKRDTEIGG 301
```

```
Qy        318 ASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVSLARLEAEI 377
              ||:|:||   : || ||   | || :||:||      :|:| | | | |  |||||| |
Db        302 HLIKEGDMVLAFVASANRDEAKFDRPHMFDIRRHPNPHIAFGHGIHFCLGAPLARLEANI 361

Qy        378 AVGTIFRRFPEMK-LKETPV 396
              |: ::   || |: :  ||:
Db        362 ALTSLISAFPHMECVSITPI 381
```

RESULT 5
ABP57702

```
ID   ABP57702 standard; Protein; 399 AA.
XX
AC   ABP57702;
XX
DT   22-JAN-2003  (first entry)
XX
DE   Saccharopolyspora ORF LVII butenyl-spinosyn biosynthetic gene product.
XX
KW   Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;
KW   metabolite; spinosyn.
XX
OS   Saccharopolyspora sp..
XX
PN   WO200279477-A2.
XX
PD   10-OCT-2002.
XX
PF   28-MAR-2002; 2002WO-US09968.
XX
PR   30-MAR-2001; 2001US-280175P.
XX
PA   (DOWC ) DOW AGROSCIENCES LLC.
XX
PI   Hahn DR,  Jackson JD,  Bullard BS,  Gustafson GD,  Waldron C;
PI   Mitchell JC;
XX
DR   WPI; 2003-058434/05.
DR   N-PSDB; ABV75558.
XX
PT   New butenyl-spinosyn biosynthetic genes, useful for increasing the
PT   production of butenyl-spinosyn insecticidal macrolides, or for changing
PT   the metabolites or products produced by spinosyn-producing
PT   microorganisms  -
XX
PS   Claim 1; Page 209-210; 218pp; English.
XX
CC   The invention relates to a novel DNA molecule comprising a DNA sequence
CC   that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
CC   PKS (polyketide synthetase) domain, or a spinosyn PKS module. The
CC   butenyl-spinosyn biosynthetic genes are useful for increasing the
CC   production of butenyl-spinosyn insecticidal macrolides. The genes are
CC   also useful for changing the metabolites or products produced by
CC   spinosyn-producing microorganisms. The present sequence represents one of
CC   the butenyl-spinosyn biosynthetic polypeptides of the invention.
XX
SQ   Sequence   399 AA;

  Query Match             24.3%;  Score 516;  DB 24;  Length 399;
  Best Local Similarity   33.4%;  Pred. No. 2.7e-41;
  Matches  127;  Conservative   81;  Mismatches  136;  Indels   36;  Gaps   10;

Qy         28 PFPAIER---------LREATPIF-----YWDEGRSWVLTRYHDVSAVFRDERFAVSREE 73
              || |:|          ||| |:        |   :|  |::|||  |:  |  | ||:
Db         11 PFGEIDRLLLDDRYAVLREGEPVSKIRLPYGGDG--WLVTRYADIKTVLGDPRFS----- 63

Qy         74 WESSAEYSSAIPELSD--MKKYGLFGLPPEDHARVRKLVNPSFTSRAIDLLRAEIQRTVD 131
                ::| : :|       :::: |  : | :| |:||||  :||:| :: ||   |: ||
Db         64 --AAAILNRDVPRGFPLILREHSLGTMDPPEHTRLRKLVGKAFTARRVEQLRPRTQQLVD 121

Qy        132 QLLDARSGQ-EEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQ 190
              |||  :        |:|   | :|:: | || :|    |:|| :          : : :
Db        122 HLLDRMAADGPPGDLVSALALPLPIKVICDLLGIPVADRERFRVWSD-----IALAITSN 176

Qy        191 VDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALVGA 250
              || :     |  :   | : ::: :|:: | | |:|::|:|| |:|::|| :|:|
Db        177 SPEEIR---ESRDQIRAYIGELVQQRKKMPTE-DLLSVLVQARAEGAQLSEEEIVVTGAG 232

Qy        251 IIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI-LRIGTVRFA 309
              .:: || :||    ||       |||   |: |: : |:| |:   |::|:||:  :      | | |
Db        233 LLIAGFETTANHIANFTFNLLTHPDQLDKLIADPELVPRAVEELLRYTPLGATPGFPRIA 292

Qy        310 RQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYGRGPHVCPGVS 369
              :|||  | ||::|: ||  | || ||  ||  ||  |: |:  : :| | ||| | |
Db        293 TEDLELGGVSIRRGDAVFFEIASANRDSAVFDGPDELDLAREHNSHMALGHGPHYCIGAQ 352

Qy        370 LARLEAEIAVGTIFRRFPEM 389
              |||:| ::|:||: :|||::
Db        353 LARMELQVAIGTLIKRFPQL 372
```

RESULT 6
AAE34729

```
ID   AAE34729 standard; Protein; 475 AA.
XX
AC   AAE34729;
XX
DT   14-MAY-2003  (first entry)
XX
DE   Streptomyces albofaciens ema8 protein.
XX
KW   P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW   emamectin; insecticide; ema8 protein.
XX
OS   Streptomyces albofaciens.
XX
PN   WO200292801-A2.
XX
PD   21-NOV-2002.
XX
PF   15-MAY-2002; 2002WO-EP05363.
XX
PR   16-MAY-2001; 2001US-291149P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Molnar I,  Ligon JM,  Zirkle RE,  Hammer PE,  Hill DS,  Pachlatko JP;
PI   Buckel TG;
XX
DR   WPI; 2003-140280/13.
DR   N-PSDB; AAD53024.
XX
PT   Novel polypeptide that exhibits enzymatic activity of P450
PT   monooxygenase and capable of regioselectively oxidizing alcohol of
PT   avermectin useful for making emamectin from avermectin  -
XX
PS   Claim 17; Page 116-117; 157pp; English.
XX
CC   The present invention relates to novel proteins that exhibit an enzymatic
CC   activity of P450 monooxygenase and capable of regioselectively oxidising
CC   the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC   The invention also relates to ferrodoxins and ferrodoxin reductases that
CC   are active with the P450 monooxygenases. Sequences of the invention are
CC   useful for producing 4-keto-avermectin from avermectin, which is useful
CC   for producing emamectin. Emamectin is useful as an insecticide. The
CC   present sequence is Streptomyces albofaciens ema8 protein.
XX
SQ   Sequence   475 AA;

  Query Match             23.1%;  Score 491;  DB 24;  Length 475;
  Best Local Similarity   29.6%;  Pred. No. 9.7e-39;
  Matches  129;  Conservative   83;  Mismatches  156;  Indels   68;  Gaps   11;

Qy          6 ANQSETKPAFDFKPFAPGYA---------------EDPFPAIERLREATPIF---YWDEG 47
              |:   |||   :   ||                 ::|:      |||  |:     : |:
Db         34 ASPDTTSPATTDRTTLPSYVGLHPGEPNLMEPELLDNPYTGYGTLREQAPLVRARFIDDS 93
```

```
Qy          48 RSWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAIPELS-----DMKKYGLFGLP--- 99
Db          94 PIWLVTRFDVVREVMRDQRF----------VNNPTLVPGIGADQDPRARLIELFGIPEDL 143

Qy         100 -----------PEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEE---FD 144
Db         144 APYLTDTILTSDPPDHTRLRRLVSRAFTARRIQDLRPRVERITDELLARLPDHAEDGVVD 203

Qy         145 VVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLVASVTE 204
Db         204 LVEHFAYPLPITVICELVGIDEEDRALWRRFGADLA-----SLNP------KRIGATMPE 252

Qy         205 GLALLHDVLDERRRNPLENDVLTMLLQA-EADGSRLSTKELVALVGAIIAAGTDTTIYLI 263
Db         253 MIAHIHEVIDERRAD-LRDDLLSGLIRAQDDDGGRLSDVEMVTLVLTLVLAGHETTAHLI 311

Qy         264 AFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEYCGASIKKG 323
Db         312 SNGTLALLTHPDQRRLIDEDPALLPRAVHELMRWCGPIQATQLRYAMEDTEVAGVQVRQG 371

Qy         324 EMVFLLIPSALRDGTVFSRPDVFDVRRDTGA----SLAYGRGPHVCPGVSLARLEAEIAV 379
Db         372 EALMFSLVAANHDPRHYTGPERLDLTRQPAGRAEDHVGFGHGMHYCLGASLARQEAEVAY 431

Qy         380 GTIFRRFPEMKLKETP 395
Db         432 GKLLTRYPDLALALTP 447
```

RESULT 7
AAE34732
ID AAE34732 standard; Protein; 475 AA.
XX
AC AAE34732;
XX
DT 14-MAY-2003 (first entry)
XX
DE Streptomyces rimosus emall protein.
XX
KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW emamectin; insecticide; emall protein.
XX
OS Streptomyces rimosus.
XX
PN WO200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP05363.
XX
PR 16-MAY-2001; 2001US-291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
DR WPI; 2003-140280/13.
DR N-PSDB; AAD53027.
XX
PT Novel polypeptide that exhibits enzymatic activity of P450
PT monooxygenase and capable of regioselectively oxidizing alcohol of
PT avermectin useful for making emamectin from avermectin -
XX
PS Claim 17; Page 121-122; 157pp; English.
XX
CC The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferrodoxins and ferrodoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces rimosus emall protein.
XX
SQ Sequence 475 AA;

```
  Query Match             23.0%;  Score 488.5;  DB 24;  Length 475;
  Best Local Similarity   29.7%;  Pred. No. 1.7e-38;
  Matches  131;  Conservative   83;  Mismatches  158;  Indels   69;  Gaps   12;

Qy           2 TQEQANQSETK-PAFDFKPFAPGYA---------------EDPFPAIERLREATPIF--- 42
Db          29 TPSAATSSDTTYPATTDRTTLPSYVGLHPGEPNLMEPELLDNPYTGYGTLREQAPLVRAR 88

Qy          43 YWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSAEYSSAIPELSDMKK-----YGLFG 97
Db          89 FIDDSPIWLVTRFDVVREVMRDQRF----------VNNPTLVPGIGADKDPRARLIELFG 138

Qy          98 LP---------------PEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEE 142
Db         139 IPEDLTPYLADTILTSDPPDHTRLRRLVSRAFTARRIQDLRPRVEQITDALLERLPDHAE 198

Qy         143 ---FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLV 199
Db         199 DGVVDLVEHFAYPLPITVICELVGIDEEDRTLWRRFGADLA-----SLNP------KRIG 247

Qy         200 ASVTEGLALLHDVLDERRRNPLENDVLTMLLQA-EADGSRLSTKELVALVGAIIAAGTDT 258
Db         248 ATMPEMIAHIHEVIDERRA-ALRDDLLSGLIRAQDDDGGRLSDVEMVTLVLTLVLAGHET 306

Qy         259 TIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEYCGA 318
Db         307 TAHLISNGTLALLTHPDQRRLIDEDPALLPRAVHELMRWCGPIQATQLRYAMEDTEVAGV 366

Qy         319 SIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGA----SLAYGRGPHVCPGVSLARLE 374
Db         367 QVRQGEALMFSLVAANHDPRHYTGPERLDLTRQPAGRAEDHVGFGHGMHYCLGASLARQE 426

Qy         375 AEIAVGTIFRRFPEMKLKETP 395
Db         427 AEVAYGKLLTRYPDLELALTP 447
```

RESULT 8
AAE34724
ID AAE34724 standard; Protein; 470 AA.
XX
AC AAE34724;
XX
DT 14-MAY-2003 (first entry)
XX
DE Streptomyces rimosus ema3 protein.
XX
KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW emamectin; insecticide; ema3 protein.
XX
OS Streptomyces rimosus.
XX
PN WO200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP05363.
XX
PR 16-MAY-2001; 2001US-291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
DR WPI; 2003-140280/13.
DR N-PSDB; AAD53019.
XX

```
PT   Novel polypeptide that exhibits enzymatic activity of P450
PT   monooxygenase and capable of regioselectively oxidizing alcohol of
PT   avermectin useful for making emamectin from avermectin  -
XX
PS   Claim 17; Page 107-108; 157pp; English.
XX
CC   The present invention relates to novel proteins that exhibit an enzymatic
CC   activity of P450 monooxygenase and capable of regioselectively oxidising
CC   the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC   The invention also relates to ferrodoxins and ferrodoxin reductases that
CC   are active with the P450 monooxygenases. Sequences of the invention are
CC   useful for producing 4-keto-avermectin from avermectin, which is useful
CC   for producing emamectin. Emamectin is useful as an insecticide. The
CC   present sequence is Streptomyces rimosus ema3 protein.
XX
SQ   Sequence   470 AA;

  Query Match             22.9%;  Score 486.5;  DB 24;  Length 470;
  Best Local Similarity   30.6%;  Pred. No. 2.6e-38;
  Matches  124;  Conservative   81;  Mismatches  147;  Indels   53;  Gaps   10;

Qy         22 PGYAEDPFPAIERLREATPIF---YWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSA 78
              |   |:|:       |||  |:      : |:    |::||:  |   | ||:||
Db         60 PELLENPYTGYGTLREQAPLVRARFIDDSPIWLVTRFDVVREVMRDQRF----------V 109

Qy         79 EYSSAIPELSDMKK-----YGLFGLP---------------PEDHARVRKLVNPSFTSRA 118
                 : :| :    |        |||:|                 | || |:|:||: :||:|
Db        110 NNPTLVPGIGADKDPRARLIELFGIPEDLAPYLTDNILTSDPPDHTRLRRLVSRAFTARR 169

Qy        119 IDLLRAEIQRTVDQLLDARSGQEE---FDVVRDYAEGIPMRAISALLKVPAECDEKFRRF 175
              |  ||  ::|  |:||:          |     |:|  :|  :|:  |  |: :      :||
Db        170 IQDLRPRVERITDELLERLPDHAEDGVVDLVEHFAYPLPITVICELVGIDEEDRALWRRF 229

Qy        176 GSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQA-EA 234
              |:  |        | |         |  : |:: |  :: :|:::||||    :|:|:  |::| :
Db        230 GADLA-----SLNP------KRIGATMPEMISHIHELIDERRA-ALRDDLLSGLIRAQDD 277

Qy        235 DGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEV 294
              || ||| |:| ||  :: || :|| :||:  ||| |:   |:  :| .|:  |: |:
Db        278 DGGRLSDVEMVTLVLTLVLAGHETTAHLISNGTLALLTHPDQRRLIDEDPALLPRAVHEL 337

Qy        295 LRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGA 354
              :|:  ::  :|:|:| :| |  |  :::|| :   : :|  |   :: |:  |: |
Db        338 MRWCGPIQATQLRYALEDTEVAGVQVRQGEALMFSLVAANHDPRHYTGPERLDLTRQPAG 397

Qy        355 ----SLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP 395
                   : :| | | | | |||| |||:| | :  |:|:: |  ||
Db        398 RAEDHVGFGHGMHYCLGASLARQEAEVAYGKLLTRYPDLALALTP 442

RESULT 9
AAE34736
ID   AAE34736 standard; Protein; 430 AA.
XX
AC   AAE34736;
XX
DT   14-MAY-2003  (first entry)
XX
DE   Streptomyces lydicus ema15 protein.
XX
KW   P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW   emamectin; insecticide; ema15 protein.
XX
OS   Streptomyces lydicus.
XX
PN   WO200292801-A2.
XX
PD   21-NOV-2002.
XX
PF   15-MAY-2002; 2002WO-EP05363.
XX
PR   16-MAY-2001; 2001US-291149P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Molnar I,  Ligon JM,  Zirkle RE,  Hammer PE,  Hill DS,  Pachlatko JP;
PI   Buckel TG;
XX
DR   WPI; 2003-140280/13.
DR   N-PSDB; AAD53031.
XX
PT   Novel polypeptide that exhibits enzymatic activity of P450
PT   monooxygenase and capable of regioselectively oxidizing alcohol of
PT   avermectin useful for making emamectin from avermectin  -
XX
PS   Claim 17; Page 128-129; 157pp; English.
XX
CC   The present invention relates to novel proteins that exhibit an enzymatic
CC   activity of P450 monooxygenase and capable of regioselectively oxidising
CC   the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC   The invention also relates to ferrodoxins and ferrodoxin reductases that
CC   are active with the P450 monooxygenases. Sequences of the invention are
CC   useful for producing 4-keto-avermectin from avermectin, which is useful
CC   for producing emamectin. Emamectin is useful as an insecticide. The
CC   present sequence is Streptomyces lydicus ema15 protein.
XX
SQ   Sequence   430 AA;

  Query Match             22.9%;  Score 485.5;  DB 24;  Length 430;
  Best Local Similarity   31.4%;  Pred. No. 2.9e-38;
  Matches  129;  Conservative   81;  Mismatches  150;  Indels   51;  Gaps   13;

Qy         22 PGYAEDPFPAIERLREATPIF---YWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSA 78
              |     |||     ||| |:      : |:    | :||: :|  | ||:||    |  |||
Db         25 PALIGDPFAGYGALREQGPVVRGRFVDDSPVWFVTRFEEVREVLRDQRF---RNNPVSSA 81

Qy         79 EYSSAIPELSDMKK-YGLFGLPPE---------------DHARVRKLVNPSFTSRAIDLL 122
                 : :    : |                 || |:|:||: :||:|  |
Db         82 --PDADPEDTPLSRLMDMMGFPEHLRVYLLGSILNNDAPDHTRLRRLVSRAFTARKITDL 139

Qy        123 RAEIQRTVDQLLDARSGQEE---FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179
              |  : :  |:||       |  |::: :|  :|:  |  |: :| |    ::| :|:
Db        140 RPRVAQIADELLARLPEHAEDGVVDLIQHFAYPLPITVICELVGIPEEDRPQWRTWGADL 199

Qy        180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQA-EADGSR 238
                   | | |  |  :::  | :         :|:::  |||  |:|:|: |:: :  ||||
Db        200 -----VSLQP--DRMSRSFPAMIDH----IHELIAARRR-ALTDDLLSELIRTHDDDGSR 247

Qy        239 LSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFD 298
              ||  |:|  :|   ::  || :|| :||      ||  |:  | |:|: |  |:  |: |::|:
Db        248 LSDVEMVTMVLTVVLAGHETTAHLIGNGTAALLTHPDQLRLLKDDPALLPRAVHELMRWC 307

Qy        299 NILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGA---- 354
                : :  :|:| :|:|   |  |:|:| | |:: || ||  ::  || |: |
Db        308 GPVHMTQLRYAAEDVELAGVRIRKGDAVQLILVSANRDPRHYTEPDRLDLTRHPAGHAEN 367

Qy        355 SLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMK-------LKETPVFG 398
                : :| | | | | :|| | |:|:| : | |||:          |: ||| |
Db        368 HVGFGHGAHYCLGATLAKQEGEVALGALLRHFPELSLAVAPDALERTPVPG 418

RESULT 10
AAB18654
ID   AAB18654 standard; Protein; 416 AA.
XX
AC   AAB18654;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Amino acid sequence of picromycin/methymycin cytochrome P450 hydroxylase.
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis;
KW   picromycin/methymycin cytochrome P450 hydroxylase.
```

```
XX
OS   Streptomyces venezuelae.
XX
PN   US6117659-A.
XX
PD   12-SEP-2000.
XX
PF   27-MAY-1999;   99US-0320878.
XX
PR   28-MAY-1998;   98US-0087080.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
PR   20-MAY-1999;   99US-0134990.
PR   30-APR-1997;   97US-0846247.
PR   06-MAY-1998;   98US-0073538.
PR   28-AUG-1998;   98US-0141908.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  Tang L,  McDaniel R;
XX
DR   WPI; 2000-610844/58.
XX
PT   New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT   for converting ketolides to antibiotics and as antibiotics and
PT   intermediates in the synthesis of compounds with pharmaceutical value
PT   -
XX
PS   Disclosure; Columns 47-48; 117pp; English.
XX
CC   The present sequence represents a picromycin/methymycin cytochrome P450
CC   hydroxylase (picK). The nucleotide sequence encoding it is used in the
CC   course of the invention. The specification describes a recombinant DNA
CC   compound expressing recombinant polyketide synthase genes in host cells
CC   for the production of narbonolide, narbonolide derivatives and
CC   polyketides that are useful as antibiotics and as intermediates in the
CC   synthesis of compounds with pharmaceutical value. The DNA compounds may
CC   also encode a C12-hydroxyalse (picK), desosamine biosynthesis and
CC   desosaminyl transferase enzymes (useful for conversion of ketolides to
CC   antibiotics), and the beta-glucosidase enzyme (involved in picromycin
CC   biosynthesis). These compounds are also useful for increasing the
CC   antibiotic activity of a compound relative to the unhydroxylated
CC   compound. The recombinant host cells are useful as genetic systems that
CC   allow rapid engineering of the narbonolide polyketide synthase. These
CC   would be valuable for creating novel ketolide analogs for pharmaceutical
CC   applications.
XX
SQ   Sequence   416 AA;

  Query Match             22.8%;  Score 484;  DB 21;  Length 416;
  Best Local Similarity   30.6%;  Pred. No. 3.8e-38;
  Matches 127;  Conservative   81;  Mismatches  183;  Indels   24;  Gaps    8;

Qy          5 QANQSETKPAFDFKPFAPGYAEDPPPAIERLREATPI--FYWDEG-RSWVLTRYHDVSAV 61
Db          5 QQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDRARAV 64

Qy         62 FRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDL 121
Db         65 LADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMRRVEL 118

Qy        122 LRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179
Db        119 LRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVWTDAF 178

Qy        180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRL 239
Db        179 -------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDEDGSRL 228

Qy        240 STKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDN 299
Db        229 TSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEMLRYEG 288

Qy        300 ILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYG 359
Db        289 PVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGHLAFG 348

Qy        360 RGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
Db        349 HGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403


RESULT 11
AAY67218
ID   AAY67218 standard; protein; 416 AA.
XX
AC   AAY67218;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Macrolide ring modifier picK hydroxylase amino acid sequence.
XX
KW   Narbonolide polyketide synthase; PKS; hydroxylase; picK; ketolide;
KW   antibiotic production; narbomycin; picromycin.
XX
OS   Streptomyces venezuelae.
XX
PN   WO9961599-A2.
XX
PD   02-DEC-1999.
XX
PF   27-MAY-1999;   99WO-US11814.
XX
PR   28-MAY-1998;   98US-0087080.
PR   28-AUG-1998;   98US-0141908.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  McDaniel R;  Tang L;
XX
DR   WPI; 2000-072618/06.
DR   N-PSDB; AAZ56001, AAZ56003.
XX
PT   New recombinant DNA encoding a domain of narbonolide polyketide
PT   synthase, for production of ketolide antibiotics -
XX
PS   Example 2; Page 40; 98pp; English.
XX
CC   This is the Streptomyces venezuelae hydroxylase PicK protein sequence.
CC   The picK protein is involved in macrolide ring modification. The
CC   invention relates to recombinant DNA containing a coding sequence for a
CC   narbonolide polyketide synthase (PKS). Polyketides are compounds
CC   synthesised from 2-carbon units through a series of condensations and
CC   subsequent modifications. Modular PKSs are responsible for the
CC   production of many antibiotics including picromycin. The narbonolide PKS
CC   consists of a loading module, six extender modules, and two thioester
CC   domains. Four proteins make up the narbonolide PKS (PICAI, PICAII,
CC   PICAIII and PICAIV). PICAI includes the loading module and extender
CC   modules 1 and 2, PICAII includes extender modules 3 and 4, PICAIII
CC   includes extender module 5 and PICAIV includes extender module 6 and a
CC   type II thioesterase domain. The second type II thioesterase domain is
CC   found on the PICB protein. The nucleotide sequences encoding all of these
CC   proteins can be isolated in recombinant form from the recombinant cosmid
CC   pKOS023-27 (see AAZ56001). Narbonolide is desosaminylated in S.
CC   venezuelae to yield narbomycin, the desosaminyl transferase enzyme is
CC   required for this conversion, and the desosamine biosynthetic genes are
CC   also found in cosmid pKOS023-27. The recombinant DNA of the invention is
CC   used to express, in transformed cells, narbonolide (or its derivatives)
CC   or other ketolides (particularly hybrids), which may then be converted
CC   (e.g. by other enzymes recombinantly expressed in the same hosts) to
CC   polyketide antibiotics or their intermediates. The antibiotics are useful
CC   in human or veterinary medicine.
XX
SQ   Sequence   416 AA;
```

```
  Query Match             22.8%;  Score 484;  DB 21;  Length 416;
  Best Local Similarity   30.6%;  Pred. No. 3.8e-38;
  Matches  127;  Conservative   81;  Mismatches  183;  Indels   24;  Gaps    8;

Qy          5 QANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHDVSAV 61
              |   : :  |  |          :| ||:|   |||   |        ||   |::  |    ||
Db          5 QQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDRARAV 64

Qy         62 FRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDL 121
                 | ||:    ::| :|    :: : |          : :    |  |  |:||||   ||  | ::|
Db         65 LADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMRRVEL 118

Qy        122 LRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179
              ||  :|  ||  |:||       :     |::    |   :|:  || || ||         ||  :   |
Db        119 LRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVWTDAF 178

Qy        180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRL 239
                      : |      : :| :| ::   |:  |     :| :|       |: :   ::   :: ||||
Db        179 -------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDEDGSRL 228

Qy        240 STKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDN 299
              :::||: :    ::.|| :||: |||   :   ||   |:  |    ::|:   |:   |::|:||::
Db        229 TSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEMLRYEG 288

Qy        300 ILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYG 359
                 :     | ||   : ::   |    |   |:  |  :::   ||       |    |    ||:||||      ||:|
Db        289 PVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGHLAFG 348

Qy        360 RGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
                |  |  |  |    ||||||  |||    :     |  |::  |    :|        |:    :|    |  :::|  :
Db        349 HGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403
```

RESULT 12
ABG71678
ID   ABG71678 standard; protein; 416 AA.
XX
AC   ABG71678;
XX
DT   21-JAN-2003 (first entry)
XX
DE   S. venezuelae picromycin/methymycin cytochrome P450 hydroxylase, PicK.
XX
KW   Narbonolide polyketide synthase; PKS; desosamine biosynthetic gene;
KW   desosaminyl transferase gene; beta-glucosidase gene; antibiotic;
KW   picK hydroxylase gene; C12 hydroxylase gene; narbonolide;
KW   desosaminylated polyketide; narbomycin biosynthesis; PicK;
KW   picromycin/methymycin cytochrome P450 hydroxylase;
KW   picromycin biosynthesis; enzyme.
XX
OS   Streptomyces venezuelae.
XX
PN   WO200297062-A2.
XX
PD   05-DEC-2002.
XX
PF   22-FEB-2002; 2002WO-US05642.
XX
PR   22-FEB-2001; 2001US-0793708.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX
DR   WPI; 2003-041412/03.
XX
PT   Preparation of polyketides by recombinant DNA technology, useful as
PT   antibiotics and as intermediates in the synthesis of pharmaceutical
PT   compounds -
XX
PS   Disclosure; Page 45; 127pp; English.
XX
CC   The present invention relates to recombinant DNA sequences encoding
CC   for a narbonolide polyketide synthase (PKS) domain, and methods of
CC   producing polyketides by recombinant DNA technology. The recombinant
CC   DNA sequences are derived from Streptomyces venezuelae desosamine
CC   biosynthetic, desosaminyl transferase, beta-glucosidase, or picK
CC   (C12) hydroxylase genes. The method is useful for transforming a cell
CC   with a recombinant expression vector that encodes a functional
CC   beta-glucosidase gene, and therefore for increasing the yield of a
CC   desosaminylated polyketide in a cell. The recombinant methods and
CC   materials are useful for expressing polyketides with significant
CC   antibiotic activity, derived in whole or in part from the narbonolide
CC   PKS gene, and other genes involved in narbomycin and picromycin
CC   biosynthesis in recombinant host cells. The present sequence
CC   represents S. venezuelae picromycin/methymycin cytochrome P450
CC   hydroxylase, PicK.
XX
SQ   Sequence   416 AA;

```
  Query Match             22.8%;  Score 484;  DB 24;  Length 416;
  Best Local Similarity   30.6%;  Pred. No. 3.8e-38;
  Matches  127;  Conservative   81;  Mismatches  183;  Indels   24;  Gaps    8;

Qy          5 QANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHDVSAV 61
              |   : :  |  |          :| ||:|   |||   |        ||   |::  |    ||
Db          5 QQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDRARAV 64

Qy         62 FRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAIDL 121
                 | ||:    ::| :|    :: : |          : :    |  |  |:||||   ||  | ::|
Db         65 LADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTRLRKLVAREFTMRRVEL 118

Qy        122 LRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179
              ||  :|  ||  |:||       :     |::    |   :|:  || || ||         ||  :   |
Db        119 LRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVWTDAF 178

Qy        180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRL 239
                      : |      : :| :| ::   |:  |     :| :|       |: :   ::   :: ||||
Db        179 -------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDEDGSRL 228

Qy        240 STKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDN 299
              :::||: :    :: || :||: |||   :   ||   |:  |    ::|:   |:   |::|:||::
Db        229 TSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEMLRYEG 288

Qy        300 ILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGASLAYG 359
                 :     | ||   : ::   |    |   |:  |  :::   ||       |    |    ||:||||      ||:|
Db        289 PVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGHLAFG 348

Qy        360 RGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411
                |  |  |  |    ||||||  |||    :     |  |::  |    :|        |:    :|    |  :::|  :
Db        349 HGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 403
```

RESULT 13
AAE34780
ID   AAE34780 standard; Protein; 430 AA.
XX
AC   AAE34780;
XX
DT   14-MAY-2003 (first entry)
XX
DE   Streptomyces tubercidicus ema17 protein.
XX
KW   P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW   emamectin; insecticide; ema17 protein.
XX
OS   Streptomyces tubercidicus.
XX
PN   WO200292801-A2.
XX
PD   21-NOV-2002.
XX
PF   15-MAY-2002; 2002WO-EP05363.
XX
PR   16-MAY-2001; 2001US-291149P.

```
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Molnar I,  Ligon JM,  Zirkle RE,  Hammer PE,  Hill DS,  Pachlatko JP;
PI   Buckel TG;
XX
DR   WPI; 2003-140280/13.
DR   N-PSDB; AAD53052.
XX
PT   Novel polypeptide that exhibits enzymatic activity of P450
PT   monooxygenase and capable of regioselectively oxidizing alcohol of
PT   avermectin useful for making emamectin from avermectin  -
XX
PS   Claim 17; Page 147-148; 157pp; English.
XX
CC   The present invention relates to novel proteins that exhibit an enzymatic
CC   activity of P450 monooxygenase and capable of regioselectively oxidising
CC   the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC   The invention also relates to ferrodoxins and ferrodoxin reductases that
CC   are active with the P450 monooxygenases. Sequences of the invention are
CC   useful for producing 4-keto-avermectin from avermectin, which is useful
CC   for producing emamectin. Emamectin is useful as an insecticide. The
CC   present sequence is Streptomyces tubercidicus ema17 protein.
XX
SQ   Sequence   430 AA;

  Query Match             22.6%;  Score 480.5;  DB 24;  Length 430;
  Best Local Similarity   29.1%;  Pred. No. 8.8e-38;
  Matches  130;  Conservative   89;  Mismatches  161;  Indels   67;  Gaps   12;

Qy         14 AFDFKPFA--------------PGYAEDPFPAIERLREATPIF---YWDEGRSWVLTRYH 56
                |   |||               |    |||      |||  |:     :  |:   |::||:
Db          3 AISSSPFAAHVGKHPGEPNVMDPALITDPFGGYGALREQGPVLPGRFMDDSPVWLVTRFE 62

Qy         57 DVSAVFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPE--------------- 101
              :|  | ||:||  :       :::    :| |   :   : |:|
Db         63 EVRQVLRDQRFLNN----PAASSPGHSIDESPTARLLDMMGMPEHFRPYLMGSILNNDAP 118

Qy        102 DHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLLDARSGQEE---FDVVRDYAEGIPMRAI 158
              || |:|:||: :||:| |  ||  :::  |:||       |    |::: :|  :|:  |
Db        119 DHTRLRRLVSRAFTARKITDLRPRVEQLADELLARLPEHAEDGVVDLIKHFAYPLPITVI 178

Qy        159 SALLKVPAECDEKFRRFGSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRR 218
                |: :|     ::|::|:              ||     |     |    :: |      :|::: | ||
Db        179 CELVGIPEADRPQWRKWGA--------DLVSLQPERLSTSFPAMIEH---IHELIRE-RR 226

Qy        219 NPLENDVLTMLLQA-EADGSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEAL 277
               | :|:|: |::   :  |||||  |:| :|  :| ||:| :||       || |:  |
Db        227 GALTDDLLSELIRTHDDDGSRLSDVEMVTMVLTVVLAGHETTAHLIGNGTAALLTHPDQL 286

Qy        278 ELVKAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDG 337
               ||| :| |:   |: |:||:   :::   :|:|  :|:|   |    |:||: | |:: ||  |
Db        287 RLVKDDPELLPRAVHELLRWCGPVQMTQLRYASEDVEIAGTPIRKGDAVQLILVSANFDP 346

Qy        338 TVFSRPDVFDVRRDTGA----SLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEM---- 389
               :: |:  |: |            : :| | | | | :||:  | |:| |  :| :||:
Db        347 RHYTAPERLDLTRHPAGHAENHVGFGHGMHYCLGATLAKQEGEVAFGKLFTHYPELSLAV 406

Qy        390 ---KLKETPVFGYHPAFRNIESLNVIL 413
                :|: |||     |      ::|| | |
Db        407 APDELERTPV----PGSWRLDSLPVRL 429


RESULT 14
AAY77196
ID   AAY77196 standard; Protein; 415 AA.
XX
AC   AAY77196;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide biosynthetic enzyme PikC, SEQ ID NO:39.
XX
```

```
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 103
FT                   /note= "Encoded by CCG"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   N-PSDB; AAZ87301.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 19; Page 429-430; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAY77190-Y77197 represent macrolide biosynthetic enzymes from
CC   Streptomyces venezuelae ATCC 15439, which are encoded by sequences
CC   AAZ87295-Z87302.
XX
SQ   Sequence   415 AA;

  Query Match             22.4%;  Score 476.5;  DB 21;  Length 415;
  Best Local Similarity   30.5%;  Pred. No. 2e-37;
  Matches  128;  Conservative   81;  Mismatches  185;  Indels   25;  Gaps    9;

Qy          1 MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREATPI--FYWDEG-RSWVLTRYHD 57
              | : |    : : |  |         :| ||:|    |||    |          ||    |:: |
Db          1 MRRTQQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGDEVWLVVGYDR 60

Qy         58 VSAVFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSR 117
              ||  | ||:   ::| :|   :: : |        : :     |  | | ||||   || |
```

```
Db         61 ARAVLADPRFS---KDWRNS---TTPLTEAEAALNHNMLESDPPRHTR-RKLVAREFTMR 113

Qy        118 AIDLLRAEIQRTVDQLLDARSGQEE--FDVVRDYAEGIPMRAISALLKVPAECDEKFRRF 175

Db        114 RVELLRPRVQEIVDGLVDAMLAAPDGRADLMESLAWPLPITVISELLGVPEPDRAAFRVW 173

Qy        176 GSATARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEAD 235

Db        174 TDAF-------VFPDDPAQAQTAMAEMSGYLSRL---IDSKRGQDGEDLLSALVRTSDED 223

Qy        236 GSRLSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVL 295

Db        224 GSRLTSEELLGMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEML 283

Qy        296 RFDNILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGAS 355

Db        284 RYEGPVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHRFDIRRDTAGH 343

Qy        356 LAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMKLKETP---VFGYHPAFRNIESLNV 411

Db        344 LAFGHGIHFCIGAPLARLEARIAVRALLERCPDLALDVSPGELVWYPNPMIRGLKALPI 402
```

RESULT 15
AAE34725

```
ID   AAE34725 standard; Protein; 430 AA.
XX
AC   AAE34725;
XX
DT   14-MAY-2003  (first entry)
XX
DE   Streptomyces lydicus ema4 protein.
XX
KW   P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KW   emamectin; insecticide; ema4 protein.
XX
OS   Streptomyces lydicus.
XX
PN   WO200292801-A2.
XX
PD   21-NOV-2002.
XX
PF   15-MAY-2002; 2002WO-EP05363.
XX
PR   16-MAY-2001; 2001US-291149P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Molnar I,  Ligon JM,  Zirkle RE,  Hammer PE,  Hill DS,  Pachlatko JP;
PI   Buckel TG;
XX
DR   WPI; 2003-140280/13.
DR   N-PSDB; AAD53020.
XX
PT   Novel polypeptide that exhibits enzymatic activity of P450
PT   monooxygenase and capable of regioselectively oxidizing alcohol of
PT   avermectin useful for making emamectin from avermectin  -
XX
PS   Claim 17; Page 109-110; 157pp; English.
XX
CC   The present invention relates to novel proteins that exhibit an enzymatic
CC   activity of P450 monooxygenase and capable of regioselectively oxidising
CC   the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC   The invention also relates to ferrodoxins and ferrodoxin reductases that
CC   are active with the P450 monooxygenases. Sequences of the invention are
CC   useful for producing 4-keto-avermectin from avermectin, which is useful
CC   for producing emamectin. Emamectin is useful as an insecticide. The
CC   present sequence is Streptomyces lydicus ema4 protein.
XX
SQ   Sequence   430 AA;

  Query Match             22.4%;  Score 475.5;  DB 24;  Length 430;
  Best Local Similarity   30.9%;  Pred. No. 2.7e-37;
  Matches  127;  Conservative   81;  Mismatches  152;  Indels   51;  Gaps   13;

Qy         22 PGYAEDPFPAIERLREATPIF---YWDEGRSWVLTRYHDVSAVFRDERFAVSREEWESSA 78

Db         25 PALIGDPFAGYGALREQGPVVRGRFMDDSPVWFVTRFEEVREVLRDPRF---RNNPVSAA 81

Qy         79 EYSSAIPELSDMKK-YGLFGLPPE--------------DHARVRKLVNPSFTSRAIDLL 122

Db         82 --PGAAPEDTPLSRLMDMMGFPEHLRVYLLGSILNNDAPDHTRLRRLVSRAFTARKITDL 139

Qy        123 RAEIQRTVDQLLDARSGQEE---FDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT 179

Db        140 RPRVTQIADELLARLPEHAEDGVVDLIQHFAYPLPITVICELVGIPEEDRPQWRTWGADL 199

Qy        180 ARALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQA-EADGSR 238

Db        200 -----VSLQP--DRMSRSFPAMIDH----IHELIAARRR-ALTDDLLSELIRTHDDDGSR 247

Qy        239 LSTKELVALVGAIIAAGTDTTIYLIAFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFD 298

Db        248 LSDVEMVTMVLTVVLAGHETTAHLIGNGTAALLTHPDQLRLLKDDPALLPRAVHELMRWC 307

Qy        299 NILRIGTVRFARQDLEYCGASIKKGEMVFLLIPSALRDGTVFSRPDVFDVRRDTGA---- 354

Db        308 GPVHMTQLRYAAEDVELAGVRIRTGDAVQLILVSANRDPRHYTDPDRLDLTRHPAGHAEN 367

Qy        355 SLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEMK-------LKETPVFG 398

Db        368 HVGFGHGAHYCLGATLAKQEGEVALGALLRHFPELSLAVAPEALERTPVPG 418
```

Search completed: October 2, 2003, 17:33:53
Job time : 30.237 secs